

数字化科研 e-Science 研究

孙 坦 主编

電子工業出版社·

Publishing House of Electronics Industry

北京 · BEIJING

内 容 简 介

全书共分四篇 14 章。对欧洲主要国家、澳大利亚、北美地区、亚洲地区的六百多个 e-Science 项目进行调研,并将相关数据收集到自行构建的 e-Science 信息门户中开展统计分析,综合采用比较分析法、案例调查法等多种科学合理的研究方法,从 e-Science 项目名称等三十多个角度,系统而深入地研究了当前世界绝大部分国家与地区的 e-Science 建设现状及相关问题,如 e-Science 的发展模式、技术架构和关键技术、管理和运行机制、e-Science 与文献情报工作之间的关系等。在此基础上,本书对我国发展 e-Science 给出了建设性的解决方案。

本书内容翔实、系统,深入浅出,覆盖面广,具有先进性、科学性和很高的实用价值。

未经许可,不得以任何方式复制或抄袭本书之部分或全部内容。
版权所有,侵权必究。

图书在版编目(CIP)数据

数字化科研: e-Science 研究/孙坦主编. —北京: 电子工业出版社, 2009.8
ISBN 978-7-121-09383-8

I. 数… II. 孙… III. 信息技术—应用—科学研究 IV. G312

中国版本图书馆 CIP 数据核字 (2009) 第 134455 号

责任编辑: 竺南直 特约编辑: 索蓉霞

印 刷:

装 订:

出版发行: 电子工业出版社

北京市海淀区万寿路 173 信箱 邮编 100036

开 本: 720×1 000 1/16 印张: 26.25 字数: 530 千字

印 次: 2009 年 8 月第 1 次印刷

印 数: 2 500 册 定价: 49.00 元

凡所购买电子工业出版社图书有缺损问题, 请向购买书店调换。若书店售缺, 请与本社发行部联系, 联系及邮购电话: (010) 88254888。

质量投诉请发邮件至 zlts@phei.com.cn, 盗版侵权举报请发邮件至 dbqq@phei.com.cn。

服务热线: (010) 88258888。

前 言

20 世纪下半叶以来,随着科研活动节奏的日益加快以及科学研究问题的不断复杂化,科研活动阵地从宏观和微观两个层面向更深、更广、更远的未知领域推进,并呈现出一系列的新特点:规模大、跨学科、合作性、高共享性和投资强度大等。人类开始进入“大科学”时代。然而,这一时代的到来,在预示着能够为人类带来更多、更重要科研成果的同时,也为实现上述特征的科学研究提出了更大的挑战。

面对“大科学”所带来的机遇与挑战,科学家不断地尝试着创造出更加有力的“工具”,以进一步扩展人类本身的能力,帮助人类达到原本无法完成的目标。于是,计算机、互联网、网格计算等“工具”相继面世,科学日新月异。随着 2000 年 11 月英国政府投入巨资建设的 e-Science 项目的正式启动,作为人类下一阶段的科学研究模式,e-Science 的理念逐渐为人们所熟悉。事实上,以“关键科学领域的全球性合作,以及实现这种全球性合作的下一代基础设施”为着眼点的 e-Science,不仅彻底深化了那场始于计算机网络的科学革命,为人类带来威力更强大的“工具”,而且将对人类科研本身以及支撑科研的相关事业带来全面的革新。

作为世界科学研究的一支重要力量,我国从一开始就十分关注 e-Science 的发展情况,并在“十五”期间大力资助网格(e-Science 的核心技术与基础设施)的研发与建设,然而在“十一五”期间,我国的 e-Science 研究与实践该如何发展,这仍然是一个重要的决策问题,需要有力的决策支持。因此,通过对国内外 e-Science 发展现状和战略的调研,分析 e-Science 的基本结构,包括其技术基础、关键技术、组成要素、运作体制,管理机制、保障机制,为我国 e-Science 的建设和发展提供指导性建议,显得意义重大。为此,本书将集中从以下几个方面开展研究:

1. 调研国内外 e-Science 的发展现状。广泛调查国内外在 e-Science 的研究和开发方面所开展的工作及取得的成果,总结国内外这一领域的发展现状。分析其研究的侧重点、特点、研究方向、取得的成果、实际价值、研究中遇到的问题及解决的方法等。

2. 分析典型的 e-Science 项目。分析国内外典型 e-Science 战略规划与计划,如英国的 e-Science 计划、荷兰 e-Science 园,探寻其设计理念、建设模式、关键技术、已有成果、战略规划等。

3. 研究 e-Science 的技术架构和关键技术。探索支持 e-Science 的各种技术基础,如网格技术、Web Service 技术、网格中间件技术、数据挖掘、知识发现等技术,分析相关技术的发展现状及趋势,研究相关技术在 e-Science 环境中的作用以及如何支持 e-Science 的运行,同时进一步分析 e-Science 的发展需要哪些关键技术提供支持。

4. 分析 e-Science 环境的组成要素及各要素之间的关系。探讨组成 e-Science 环境的各种要素及各要素之间的关系。分析 e-Science 如何有效地支持科学研究,剖析 e-Science 环境中支持科学研究的各个系统(如文献情报服务系统、仪器设备共享系统、科学数据管理系统等)的结构、组成、使命及彼此之间的关系。

5. 探究 e-Science 环境的管理和运行机制。探讨 e-Science 的建设需要采用什么样的管理机制;如何从宏观上规划,微观上实施。分析 e-Science 的具体运行机制,如何建立起良性的运行机制,这种运行机制需要什么样的管理政策与之相匹配。

6. 总结 e-Science 的发展模式。探讨 e-Science 的发展轨迹,分析发展过程中可能经历的若干典型阶段,各个阶段的基本形态、研究重点、主要特征以及构建 e-Science 的发展模式。

7. 探索 e-Science 与文献情报工作之间的关系。探索 e-Science 与文献情报工作的关系,e-Science 环境下文献情报工作与数字图书馆的关系,寻找文献情报工作融入 e-Science 环境的切入点及有效方式,力求使文献情报机构服务成为我国发展 e-Science 的有效支撑服务。

在细分研究内容的基础上,编写组通过调研当前国内外 e-Science 的发展现状,结合世界各国、地区 e-Science 当前的技术发展特点,深入分析了构建 e-Science 环境的各项关键技术。在此基础上,利用调研数据进行分析 and 总结,完成了 e-Science 的规划与战略调研。通过对中国 e-Science 项目的分析,从当前中国文献情报机构的发展现状与未来发展趋势出发,结合已完成的多维度的分析调研,本书试图为我国文献情报事业在 e-Science 环境下如何获得更为宽广的发展空间提出战略层面的建议。

本书是集体智慧的结晶。在编写本书的过程中,国家科技文献图书中心袁海波主任、国家科技文献图书中心沈仲祺副主任给予了大力支持。在撰写的过程中,编写组采用分工协作的策略。其中,参与第一篇写作的作者是:孙坦,黄国彬,刘细文,徐坦,虞惠达;参与第二篇写作的作者是:张智雄,常唯,李飒,曲云鹏;参与第三篇写作的作者是:孙坦,刘细文,张智雄,黄国彬,周静怡,徐坦,虞惠达;参与第四篇写作的作者是:孙坦,黄国彬,常唯。同时,硕士研究生黄飞燕、余倩、姜晓曦为本书的图表制作提供了大力帮助。本书最后由孙坦统一审稿。

e-Science 是一项新生事物,而且发展迅速。囿于本书编写组的学识与能力,加之编写时间紧迫,书中恐有诸多疵误,恳请学界同仁不吝指正。

孙 坦

2009-3-25 于北京

《数字化科研——e-Science 研究》

编撰人员名单

主 编：孙 坦

副主编：黄国彬 张智雄 刘细文

编写组成员

孙 坦	刘细文	张智雄	黄国彬	常 唯
曲云鹏	李 飒	周静怡	徐 坦	虞惠达

目 录

第一篇 发展现状篇

第 1 章 e-Science 概述	3
1.1 e-Science 出现的必然性分析	3
1.1.1 生产力发展的需要	3
1.1.2 共享理念的推动	3
1.1.3 网格技术的支撑	4
1.2 e-Science 的本质	5
1.3 e-Science 研究与建设的主要内容	6
1.4 e-Science 的发展特点	7
1.5 e-Science 环境下科研模式与知识流	8
1.5.1 e-Science 环境下的科研模式	8
1.5.2 e-Science 环境下的科学研究知识流	9
第 2 章 英国 e-Science 发展现状分析	11
2.1 英国 e-Science 建设背景	11
2.1.1 英国 e-Science 发展原因	11
2.1.2 英国 e-Science 发展目标	12
2.1.3 工作原理	12
2.1.4 英国 e-Science 建设的经费分配	13
2.2 英国 e-Science 承担主体发展目标分析	14
2.2.1 七大研究理事会的 e-Science 发展目标	14
2.2.2 国家 e-Science 研究中心的 e-Science 发展目标	16
2.2.3 十大地区中心的 e-Science 发展目标	17
2.2.4 七大 e-Science 优秀研究中心	19
2.3 英国 e-Science 技术框架与技术模块	21
2.3.1 英国 e-Science 组织框架	21
2.3.2 三十一个技术组件	22
2.3.3 四层架构的建设模式	22
2.3.4 二十六项技术模块	23
2.4 英国 e-Science 当前进展	24
2.4.1 英国 e-Science 核心计划	24

2.4.2	英国国家 e-Science 中心参与的 e-Science 项目	25
2.4.3	英国其他 e-Science 中心参与的 e-Science 项目	29
2.4.4	典型 e-Science 项目进展	31
2.5	小结	32
第 3 章	美国 e-Science 发展现状分析	37
3.1	美国 e-Science 历史背景	37
3.1.1	技术背景分析	37
3.1.2	应用历史分析	38
3.2	美国 e-Science 项目分析	38
3.2.1	项目选取方法说明	38
3.2.2	资助结构分析	39
3.2.3	主要资助机构和资助计划介绍	42
3.2.4	项目承担机构的构成分析	46
3.2.5	项目组成分析	49
3.3	美国 e-Science 技术框架与组件	51
3.3.1	网格框架	51
3.3.2	软件组件	53
3.4	美国 e-Science 当前进展	55
3.5	小结	57
第 4 章	欧洲 e-Science 发展现状分析	58
4.1	欧洲 e-Science 历史背景	58
4.2	欧洲 e-Science 承担主体分析	61
4.3	欧洲 e-Science 技术框架与技术模块	66
4.3.1	网格技术层次分析	66
4.3.2	网格技术开发领域分析	68
4.3.3	网格技术应用类型分析	68
4.4	欧洲 e-Science 当前进展	69
4.4.1	项目技术成就分析	69
4.4.2	项目开发出的组件分析	70
4.4.3	项目应用成就分析	71
4.5	小结	71
第 5 章	亚洲 e-Science 发展现状分析	73
5.1	中国大陆 e-Science 发展情况	73

5.1.1	中国国家网格 (CNGrid)	74
5.1.2	上海信息网格 (ShanghaiGrid)	77
5.1.3	中国教育科研网格 (ChinaGrid)	80
5.1.4	国家自然科学基金委网格建设项目 (CROWN)	83
5.1.5	863 空间信息网格	88
5.2	中国台湾 e-Science 发展情况	91
5.3	韩国 e-Science 历史与现状	92
5.4	日本 GRID 历史与环境	92
5.5	小结	94
5.5.1	各个项目承担主体分析	94
5.5.2	亚洲 e-Science 的技术框架与技术模块分析	95
第二篇 关键技术篇		
第 6 章	基于网格的资源和服务共享技术	99
6.1	网格概述	99
6.1.1	网格的概念和主要类型	99
6.1.2	网格在科学研究中的应用类型	101
6.1.3	网格技术的研究重点	102
6.1.4	网格的基本组件和功能	104
6.2	网格的主要技术标准	106
6.2.1	开放网格标准体系 (OGSA)	107
6.2.2	开放网格基础框架	117
6.2.3	数据存取与集成标准 (OGSI-DAI)	118
6.2.4	Web 服务资源框架	120
6.3	与网格相关的共享和集成技术	122
6.3.1	Web 服务技术	122
6.3.2	网格门户技术	130
6.4	Globus	137
6.4.1	GT3 (Globus Toolkit3)	137
6.4.2	GT4 (Globus Toolkit4)	140
6.5	发展趋势	153
第 7 章	科研数据的采集、管理、保存与分析技术	155
7.1	数据采集技术	155
7.2	数据集成技术	158

7.2.1	新型 OGSA-DAI 框架	158
7.2.2	BRIDGES——基于 OGSA-DAI 进行信息集成的实例	164
7.2.3	BDWorld——大规模数据抽取整合实例	166
7.2.4	eSDO——大数据量整合实例	169
7.3	数据存储和管理技术	171
7.3.1	Geodise: 基于网格的工程数据管理	171
7.3.2	BioSimGrid: 基于网格的分布式数据技术	172
7.3.3	MySpace: 虚拟观测台的数据管理技术	176
7.3.4	Data Portal: 使用门户进行数据管理的技术	178
7.3.5	SRB: 存储资源代理技术	181
7.4	元数据管理技术	184
7.5	数据保存技术	187
7.5.1	数据保存的三个概念	187
7.5.2	OAIS 参考模型——数字保存的基础	187
7.5.3	DCC 数字保存的三个阶段	193
7.6	数据分析处理技术	196
7.6.1	两种数据分析技术	196
7.6.2	e-Science 环境下的文本挖掘技术	199
7.6.3	基于网格的知识发现服务技术	201
第 8 章	研究对象的建模和仿真技术	203
8.1	可视化技术	203
8.1.1	基于网格的可视化技术框架	203
8.1.2	GViz 分析	206
8.1.3	e-Demand 分析	208
8.1.4	飞行器中的电磁散射	209
8.2	虚拟观测台技术	210
8.2.1	VO 概念	211
8.2.2	AstroGrid 项目	211
8.3	计算机动画技术	213
8.3.1	计算机动画技术研究	214
8.3.2	The PGPGGrid Project 项目分析	216
第 9 章	虚拟研究团队的组建和协同技术	219
9.1	虚拟组织技术	219
9.1.1	虚拟组织的概念	219

9.1.2	DAME 的动态虚拟组织技术	221
9.1.3	ICENI: 虚拟组织管理门户	222
9.1.4	eMineral Project: 计算门户框架	224
9.2	虚拟研究环境技术	226
9.2.1	概念	226
9.2.2	虚拟研究环境的技术基础	227
9.2.3	SAKAI: VO 中间件	230
9.2.4	IB VRE: 对研究过程的支持	231
9.3	学术交流技术	233
9.4	协作工具	236
9.4.1	基于网络的协作工具——Access Grid	236
9.4.2	虚拟组织建设——eMinerals	239
9.4.3	e-Science 协作调动空间——CoAKTinG	240
9.5	问题求解环境	242
9.5.1	协作医学问题求解——MIAKT	243
9.5.2	分布式飞机维护环境——DAME	246

第三篇 规划发展篇

第 10 章 e-Science 的规划与管理

10.1	e-Science 已经成为发达国家科研模式创新的方向	251
10.2	网格技术成为 e-Science 核心技术	253
10.3	基本形成统一的 e-Science 技术体系	256
10.4	大规模的合作成为各国 e-Science 建设的主要方式	258
10.4.1	跨国家合作是各国 e-Science 建设的特征之一	258
10.4.2	高校、科研机构通力合作是 e-Science 建设的又一特征	261
10.4.3	项目承担机构在 e-Science 合作建设中角色定位各不相同	269
10.5	各国 e-Science 建设与学科领域、具体应用紧密结合	271
10.6	各国 e-Science 规划实施各有特色	275
10.6.1	e-Science 规划布局模式不同	275
10.6.2	e-Science 建设具有明显的技术层次性	276
10.7	政府在 e-Science 建设的宏观规划中发挥主导作用	281
10.8	政府在 e-Science 建设的管理中发挥主导作用	283
10.8.1	以英国和欧盟为代表的集中式管理运行模式	283
10.8.2	以美国为代表的分散式管理运行模式	293
10.9	政府是 e-Science 建设的主要投资者	294

第 11 章 中国 e-Science 规划与建设分析	295
11.1 中国 e-Science 规划与建设的主要特征	295
11.2 中国 e-Science 规划与建设与国外的差距分析	300
11.2.1 从网格基础设施层分析我国的发展差距	300
11.2.2 从网格中间件层分析我国的发展差距	300
11.2.3 从应用开发环境与工具层分析我国的发展差距	306
11.2.4 从具体应用层分析我国的发展差距	306
11.3 中国发展 e-Science 策略分析	307

第四篇 支撑服务篇

第 12 章 e-Science 环境下文献情报机构发展分析	317
12.1 e-Science 对文献情报机构工作环境的影响	317
12.1.1 e-Science 环境下科学研究的过程分析	317
12.1.2 e-Science 给科学研究带来的新变化	318
12.2 e-Science 环境下文献情报机构的服务对象分析	319
12.2.1 用户类型分析	319
12.2.2 用户信息需求的变化	320
12.3 e-Science 环境下的文献情报服务	321
12.3.1 e-Science 环境下的文献情报服务及其定位	321
12.3.2 e-Science 环境下文献情报服务的作用	323
12.3.3 e-Science 环境下文献情报服务工作的指导原则	325
12.4 e-Science 环境对数字图书馆的影响	326
12.4.1 e-Science 对数字图书馆的积极影响——以 SRB 为例	326
12.4.2 积极应对 e-Science 的数字图书馆	328
12.4.3 服务 e-Science 的数字图书馆	328
第 13 章 e-Science 环境下文献情报机构的服务模式	330
13.1 e-Science 环境下文献情报机构的服务模式	330
13.1.1 融入知识创造过程的知识服务	330
13.1.2 隐性知识的管理和利用	333
13.1.3 分布式资源体系的建设和管理	335
13.1.4 科学数据的管理	337
13.1.5 构建开放的数字化网络化学术交流体系	339
13.1.6 以网络信息意识的强化为重点的用户信息素养的培育	342
13.2 e-Science 环境下文献情报机构服务的实现模式	344

13.2.1	服务理念升华——从“以需求拉动服务”到“以服务激发需求”	344
13.2.2	组织模式的分化——嵌入式和支持中心的双层模式	345
13.2.3	服务手段的改进——实现零障碍服务	348
13.2.4	人员构成的虚拟化——虚拟动态团队的人员构成机制	351
13.2.5	e-Science 环境下文献情报服务的保障机制	352
第 14 章	e-Science 环境下的数字图书馆	354
14.1	e-Science 环境下数字图书馆范式的演变	354
14.1.1	聚焦于数字化资源的数字图书馆范式	354
14.1.2	强调集成化服务的数字图书馆范式	356
14.1.3	虚拟数字图书馆——e-Science 时代的数字图书馆范式	359
14.2	e-Science 环境下数字图书馆的功能框架	366
14.3	演变后的功能特点分析	367
参考文献		371
附录 A	美国 e-Science 相关项目列表	376
附录 B	欧盟第五框架计划和第六框架计划资助的项目	380
附录 C	欧盟第五框架计划下网格项目研究网格技术层次	382
附录 D	欧盟第五框架下网格项目开发的组件	383
附录 E	中国 e-Science 相关项目列表	385
附录 F	调研 e-Science 项目技术研究内容	386
附录 G	参与国际合作项目 3 个以上的 29 个国家之间的合作矩阵	395
附录 H	调研 e-Science 项目的主要学科及应用领域	396
附录 I	英国 e-Science 评估指标	398
附录 J	CORE、七大研究理事会 e-Science 项目评估指标	403
附录 K	本书所用缩略语和中英文对照表	405

第一篇 发展现状篇

本书编写组首先对英国、欧洲地区、澳大利亚、北美地区、亚洲地区的 679 个 e-Science 项目进行调研，并将相关数据收集到项目组内部使用的 e-Science 研究和实践现状的信息门户中，其中设置的属性类型包括：项目名称、项目名称缩写、项目目标、项目网址、项目所属地区、资金来源、资金年限、资金数量、项目启动时间、项目截至时间、项目简介、项目详介、承担机构名称、承担机构所属国家、承担机构性质、承担机构所属学科、承担机构研究领域、项目应用领域、项目类型、项目参与总人数、负责人姓名、负责人学科背景、项目成果形式、项目是否国际合作、项目功能、项目开发语言、项目直接架构平台、项目基础技术平台、项目是否开源、项目使用的工具包、项目开发的工具包、项目开发的工具包下载地址、项目 OGSA 技术层面、项目母项目、项目所属类别。在项目调研的基础上，选取英国、美国、欧洲和亚洲等国家和地区的 e-Science 发展现状为研究重点，完成了世界当前 e-Science 的发展现状调研（截止于 2006 月 12 月）。调研结论将主要分布在发展现状篇和规划发展篇中。

第 1 章 e-Science概述

1.1 e-Science出现的必然性分析

1.1.1 生产力发展的需要

科学研究是人类认识自然、改造自然的一项重要活动。科学研究的结果在推动社会发展、人类进步的过程中有着非同寻常的作用。在 19 世纪、20 世纪，人类社会的科学研究活动取得了巨大成就。但是，科研活动的开放度明显不够，在科研过程中，科学家之间的信息交流极其匮乏。在这种情况下，不但无法共享彼此的成果、传播科研经验，更为严重的是出现了许多重复性劳动。比较典型的是，曾经有许多不同的人几乎在同一时间内各自独立的做出了相同的成果。同时，对于一些高难度、大规模的研究，由于缺乏相关的资源或条件，依靠传统的科学研究手段往往会受到诸多制约。如需要有大規模军队参与才能完成的项目，或者是需要在一些人类根本无法进入的环境中进行实验，包括超高温、超低温环境。

科学问题的空前复杂化是科学技术发展进程中面临的又一个难题。科学研究对象已不再是简单的孤立系统，而是涵盖更大的范围、横跨多个学科。时至今日，科学研究的前沿正微观与宏观这两个层面上向着更为广阔的领域前进。无论是微观粒子、宏观宇宙、染色体、基因，还是航天工程、全球气候、生态环境等问题，都已经令人类的研究达到一个更为复杂深奥的程度。

1.1.2 共享理念的推动

“交流”与“共享”是科学研究永恒的主题。网络革命性地改变了科学研究交流与共享的方式，为科学研究的充分共享与交流提供了极大的发展空间，将交流、合作与共享推广到科学研究活动本身乃至贯穿于整个科学研究过程之中，标志着科学研究共享时代的全面到来——科学家群体共享的对象不仅包括传统的数据、资料、信息，更增加了科学家的智慧与劳动以及科学仪器设备的共享。

1. 科学数据共享的需要

科学数据是科学研究得以顺利进行的核心资源，已被人们视为是一种战略资源。而科学数据共享问题作为科学发展政策的研究起源于发达国家，经过多年的发展，现在已成为信息社会中科技发展的一项重大方针。随着计算机与新一代 Internet 的迅猛发展，科学研究的信息共享出现新的格局。以英国为例，它所提出了 e-Science 计

划，目标之一就是要以网格（Grid）为基础设施实现科学数据的共享。通过充分发挥高性能的网络和计算能力，科学家个体和用户在获取信息时，不需知道所获取的数据来自何方、是通过哪一台计算机的运转而得，就可得到科学研究所需的数据。当然，由于数据是在共享的理念下被构建、存储以及操作，所以，这些数据往往是跨部门、跨地区的；如果有双方或多方协议的许可，这些数据也可以是跨国的。

在信息化时代，科学数据共享是推动科学技术可持续高速度发展的驱动力，并已成为一个国家综合国力的衡量指标之一。它作为学术研究课题被提出始于 1979 年 5 月。目前，我国正在加大力度推进科学数据共享。2002 年 11 月，科技部徐冠华部长主持了以科学数据共享为主题的第 196 次香山科学会议，这在一定程度上标志着我国科学数据共享将进入一个新的发展阶段。

2. 科学仪器设备共享的推动

以往许多昂贵的科学仪器设备往往只能供部分科研单位使用，而其他科研单位只能“望洋兴叹”。需要使用该设备时，即使得到设备所属单位的许可，这些科研单位也只能到设备的实际物理存放地点进行操作。这常常给科学研究活动带来许多不便与障碍。共享的理念，就是要实现这些科学仪器设备可以跨越时间、空间被处于不同地理位置的部门进行操纵与控制，实现共享。

缺乏科学数据与科学仪器设备的共享，往往使得许多有可能提前得到突破的成果姗姗来迟。为了扭转这一层面，人们对 e-Science 的到来翘首以待。也正是因为存在这种迫切性，使得 e-Science 一经英国首先倡导，便如雨后春笋般在全世界范围内迅速蔓延。由于发达国家的信息化程度高，科研水平发达，拥有丰富的信息资源和关键的科学数据，因此，发达国家与发展中国家之间的科学数据共享尤显重要。

1.1.3 网格技术的支撑

美国信息处理学会联合会（American Federation of Information Processing Societies, AFIPS）在 1970 年从共享资源出发，把计算机网络定义为“以能够相互共享资源（硬件、软件和数据库等）的方式连接起来，并各自具备有独立功能的计算机系统的集合。”

如前文所述，当今的科学研究对象日益复杂，科学研究本身也变得更加需要合作和多学科综合。一个科学研究小组的成员可能分布在不同的研究单位、地区和国家。当前的技术，如通过 E-mail 和 Web，这些科研人员可以获得一起合作的基本机制。但是，它依然需要科研人员花费大量时间学习和掌握新增加的有关计算资源的技术细节、访问远程的应用程序或者等待远程计算资源的响应。如果能够将这些分布式的数据、计算机、传感器和其他资源连接成一个虚拟的实验室，那将会是怎样的景象呢？网格技术的目的正是通过提供协议、服务和能够实现灵活可控大规模资

源共享的软件开发包，让这一构想切实可行。

网格是伴随着互联网技术发展而迅速发展起来的。最初，它是专门针对复杂科学计算应用的一种新型计算模式。这种计算模式的特征就是要把整个网络整合成一台巨大的超级计算机。随着网格技术的发展和应用面的扩展，网格将演变为具有高性能处理、海量数据存储和大量仪器设备等特征的 21 世纪人类社会的信息处理基础设施。

一般而言，网格主要由六部分组成，即网格节点、数据库、贵重仪器、可视化设备、宽带主干网和网格软件。其中，网格节点是网格计算资源的提供者，它包括高端服务器、集群系统、MPP（Massively Parallel Processor）系统大型存储设备、数据库等；宽带网络系统内嵌在网格计算环境中，它是实现高性能通信的基础；资源管理和任务调度工具主要用于解决资源的描述、组织和管理等关键问题，属于网格计算的中间件，是网格技术得以将资源无缝提供给需求者的关键；监测工具帮助用户充分利用网格计算中的资源；可视化工具设备存在于网格的应用层，它通过友好的用户界面，将计算结果转换成直观的图形信息，帮助研究人员克服数据理解上的困难。

可以说，作为 Internet 信息技术的下一个浪潮的网格技术，将是 e-Science 得以实现的核心技术。也正是因为它的存在与迅猛发展，人们才有信心提出 e-Science 的构想。因此，在网格技术的强大支撑下，在变革科学研究模式，提高科学研究的合作、交流与共享程度，提高科学研究效率的强烈愿望的推动下，人们构造出一种全新的科研协作模式和大科学工程，这就是 e-Science。它以新一代互联网技术和网格技术为基础，通过 Internet 联合组成一个共同的虚拟研究团队，共享资源和成果，协同工作共同完成大型的现代科学研究。

1.2 e-Science的本质

尽管 e-Science 出现的历史短暂，但是，迄今为止，已经有多家研究机构以及多位研究专家从不同角度就 e-Science 的本质认识展开论述。欧洲生物信息学研究所认为，e-Science 就是依赖通信技术与计算基础设施的发展，开发出一种对科学活动的各个方面提供支持的数字技术，从而给科学家的工作提供支撑。英国著名科普杂志《新科学家》指出，e-Science 是一种日益依靠以因特网为支撑的分布式全球合作来进行的科学研究。为实现其目标，它必须处理大型数据集，并利用万亿级的计算资源以及高性能可视化。

按照Christine L. Borgman的观点，e-Science的目标是构建一种新型的科学研究模

式,这种新型科研模式具有信息密集、数据密集、分布式、协作和多领域的特征¹。其中,信息技术起到三方面作用:提高科学数据在科学研究中的重要价值,将科学数据处理作为一种新的科学投入模式;管理“数据洪流”,实现有效的数据管理和利用;促进科学信息更加方便快捷地存取。

那究竟什么是 e-Science? 可通过下面两个实例来说明。

CombeChem项目是EPSRC (Engineering and Physical Sciences Research Council, 英国工程与自然科学研究委员会)的e-Science试验项目²。它利用基于网格的信息和知识共享环境,可以集成现有数据源的结构和属性。这个基于服务的网格计算基础设施,包括实验室中的各种试验设备,通过大量的原始数据的积累、集成、分析、关系建模、仿真等过程,可以实现虚拟的科学研究。

Geodise (Grid Enabled Optimisation and Design Search, 基于网格的优化和设计查询)是一个利用分布式计算资源,实现计算流体力学方面设计优化的项目,它来自于设计优化、计算流体力学、网格计算、知识管理和知识本体的各种技术有机结合,形成一个应用于产业领域,实现问题求解的解决方案³。这一项目中,它基于网格,能够无缝地存取智能知识库、一系列的优化和检索工具、工业分析代码以及分布式的计算和数据资源。

我们认为: e-Science 就是使用计算和通信技术,通过全球性的合作来促进科学研究活动的开展。e-Science 整合了网络计算、程序设计工具、数据和可视化技术,这些使得科学家可以通过直接的仿真、实验和观察的比较,获得有关其实验研究的更深入见解,通过 e-Science 研究人员将可以访问到支持关系、概念和推论规则的知识库,从而为其获得新的见解提供便利。

1.3 e-Science研究与建设的主要内容

目前,国内外有关 e-Science 的研究,主要集中在以下四方面:利用信息技术进行研究数据的采集、获取、保存和分析问题;数字环境下被研究对象的建模和仿真问题;动态的、分布式的、虚拟的研究团队的组建和工作模式问题;网格的研究与应用以及由此引发的计算力等资源的共享等问题。

在建设方面,完善 e-Science 基础架构是世界各国家和地区现阶段的建设重点。在实施 e-Science 核心计划的过程中,英国确定了发展 e-Science 必需的三十一类通用

¹ Christine L. Borgman.The role of libraries in e-Science[EB/OL].[2009-02-17].http://www.eahil.net/conferences/helsinki_2008/www.terkko.helsinki.fi/bmf/EAHIL2008/Borgman-The-Role.pdf

² CombeChem [EB/OL].(2008-8-20),<http://www.combechem.org/index.php>

³ Geodise [EB/OL]. (2008-8-20),<http://www.geodise.org/>

组件，并辅以九十一个试验项目展开可行性研究。这三十一类组件包括：应用服务；网格计算服务；领域本体、任务本体与元数据；发布与归档服务；问题求解环境；合作工具包；数据管理服务；数据挖掘与分析服务；网格管理服务；仪器管理服务；市场经济服务；模式与程序原型；时序安排、监测与诊断服务；虚拟与沉浸环境服务；可视化服务；工作流服务；审核；认证与授权；代理；合作与远程仪器服务；协同调度；数据访问服务；数据编目与溯源；容错管理；全局事件服务；网格信息服务；全局请求；监测；服务质量的可靠性；资源访问服务；安全与隐私。

1.4 e-Science的发展特点

概括起来，e-Science 在发展过程中表现出来的特征主要包括：

（1）网格技术是核心支撑

网格技术集成了计算机科学领域中多种技术的最新成果，代表了信息技术和信息化发展的方向，尤其是在现阶段，突出体现和满足了科学研究的需要。正是通过网格才有可能将 e-Science 所需的众多技术与资源集成一体。

（2）与特定学科密切结合

目前，从各国 e-Science 项目的进展现状，尤其是英国各 e-Science 的发展现状可以发现，e-Science 的研究与建设基本上都与特定学科相结合。如 PPARC (Particle Physics and Astronomy Research Council) 在 e-Science 的研究与开展过程中，就密切结合了天文学。基本上，英国六大研究理事会也均是结合一定的学科领域来开展 e-Science 建设。

（3）开展合作始终是主线

在开展 e-Science 试验研究与初步建设的过程中，各国均非常重视开展各种各样的合作。这种合作包括跨行业间的合作：以英国为例，其 e-Science 核心计划从一开始就注重与工业界的合作，英国国家 e-Science 中心所公布的“中心工业项目”就达到 49 个。通过与工业界的合作，不仅可以利用工业界的资金与技术，更可以将工业界的研发人才吸引到 e-Science 的研究阵容中，最终有利于促进研究上的集思广益。同时也有利于研究成果迅速在工业界得到推广。同时还包括国家间的合作。例如，英国 e-Science 核心计划就强调，建设 e-Science 需要“参与到国际网格项目和活动中”。目前，中国与英国在生命科学研究领域上的 e-Science 合作进展顺利。

（4）共享一直是追求目标

从当前各国 e-Science 的建设与研究现状来看，共享一直是各国 e-Science 建设追求的目标。e-Science 环境下的共享，不仅可让研究人员在高校校园内开展共享，更可以让他们在各城市之间、各州之间、全国范围内以至全球范围内进行共享。

1.5 e-Science环境下科研模式与知识流

1.5.1 e-Science环境下的科研模式

海量数据处理、高性能计算与高速网络既是推动网格出现的三大要素，也是e-Science开展过程中遭遇的三大问题。图1-1描述了e-Science出现之前的科研模式下，科研人员对于科学数据以及与科学数据有关的各种操作，如数据的获取（大多通过实验获得数据）、计算与分析均是一种分散性的行为。为了完成某一科研任务，科研人员往往需要在这些分散性行为上耗费时间与精力，科研效率明显不高。

在e-Science环境下，科研模式将发生极大的变革。与以往的科研模式不同，科学家直接面对的将不再是各种分散的数据操作，而是通过网格技术及其相应的中间件实现程序的各种分散性操作的集成。在这种环境下，科学家只需要提交任务请求，便可以通过单一的入口，而无须考虑具体实现过程，接受集成化服务，从而大大提高了科研效率。

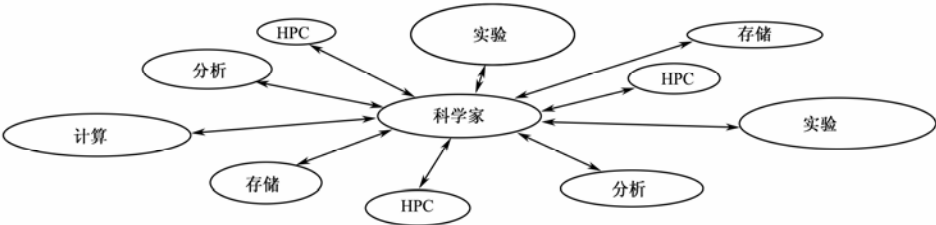


图 1-1 e-Science 出现之前的科研模式

图1-2展示了e-Science环境下新型的科研模式。该图共由三部分组成，图的两端分别为科学家、各种相互联结的操作（实验、存储、分析、计算等），图的中间部分则是可以将科学家和集成操作沟通起来的“桥梁”——网格技术及其相关的中间件实现程序。网格技术与中间件实现程序在e-Science开展的过程中担负两种角色：一是要将以往针对数据的各种分散性操作加以联结和集成；二是负责为科学家访问这种集成服务提供入口。这都要求中间件与网格必须具有高吞吐量的数据处理能力。

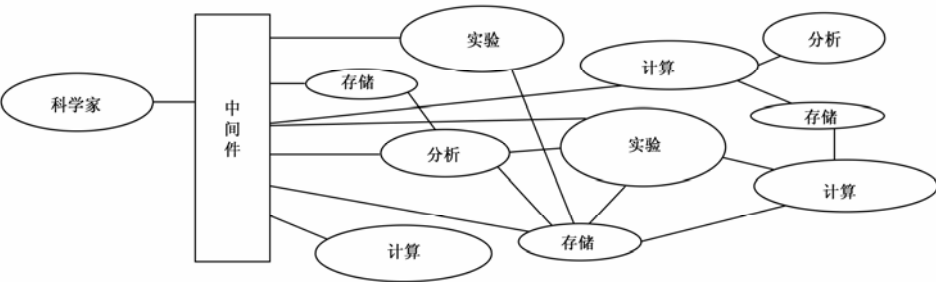


图 1-2 e-Science 环境下的科研实现模式图

数据吞吐能力上的优势使得网格可以最大限度地、健壮地、可扩展地、充分利用闲置资源；使得它可以根据科研需要，对各种操作实施定制。为完成某一任务请求，网格可以对某些操作实施灵活的时间限制；网格所面对的数据密集性，使得它在数据处理方面必须具备计算性、可评估性、可存储性与可恢复性；分布式的超级计算能力使得网格可以汇集众多 CPU 与内存，以突显合作性的方式共享资源完成各种任务。

在网格技术的推动下，科学家也因为 VO 的到来而在外延上变得更为丰富。虚拟组织是为了实现某一临时的科研目标而临时建立起来的一支虚拟团队。它具有透明地访问网格资源、不需关注资源的物理位置、只需在乎资源的属性及可获取性等特点。它是 e-Science 环境下科研活动的重要主体。网格是一种可以在动态的、涉及多个学科的虚拟组织中开展资源共享与协同工作求解的技术。由此可见，在 e-Science 科研模式下，虚拟组织是一个不容忽视的科研虚拟实体。它是科学家这一主体在 e-Science 时代的延伸。

1.5.2 e-Science环境下的科学研究知识流

在 e-Science 环境下，科学研究涉及到科学家（包括虚拟组织）、门户网站、元数据管理者、数据掌管中心（Data Curation Center）、机构知识库、同行评议等多个因素，围绕这些因素之间而产生的知识流动、数据流动便构成了 e-Science 环境下知识流轨迹。

在图 1-3 所示的科学研究知识流程中，可以发现，知识流的走势是环形的，具有可重复的周期性。在 e-Science 这种协作科研大环境下，数据或者是知识被创建完毕后，将通过机构存储或者是自我存档这一环节，由机构存储库或者个人知识存储库利用元数据技术进行组织、管理与存储。为了在不同的机构存储库或者是个人知识库之间进行互操作，元数据收割协议提供相应的协助措施。而借助门户网站或者是数据掌管中心提供的服务，科学家或者虚拟组织可以获取并收集来自机构知识库或者个人知识库的相关资料从而开展科研活动。在完成科研活动后，其科研成果将通过同行评议期刊或者是会议论文等渠道加以发表并被收集到机构知识库或者个人知识库中供后续科研活动调用。在此之后，科研人员又开始酝酿下一次科研活动，开展科研所需的知识流、数据流也将开始新的循环周期。如此反复，不断循环，通过知识不断地循环流动，最终实现科研数据、科研资料的高效利用。

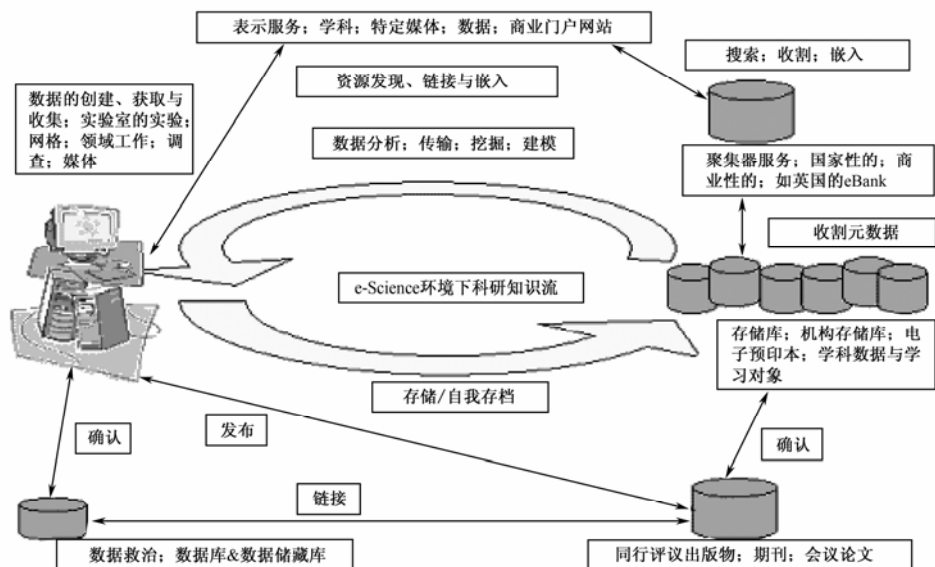


图 1-3 e-Science 环境下科学研究的知识流

第2章 英国e-Science发展现状分析

2.1 英国e-Science建设背景

2.1.1 英国e-Science发展原因

从20世纪90年代末的首次出现,到2001年英国e-Science核心计划的正式实施,无论是从理论层面,还是实践角度,英国e-Science出现的时间都较为短促。尽管如此,在短短五年内,英国已经投入大量资金,并在全英国范围内启动了三三百个以上的e-Science项目。通过深入分析可知,英国如此大力进行e-Science建设,主要基于两方面的考虑。

(1) 科研发展的需要是推动英国e-Science蓬勃发展的根本原因

众所周知,传统的科学研究都是基于实验的,或者以理论分析为主,但是这种实验的、理论的分析其实存在很多问题,比如比较封闭,没有模拟,没有仿真的手段和方法,因而造成科研周期较长,甚至成本很高,还有些预测性的实验根本没法开展。

人类进入二十世纪末二十一世纪初,科学研究工作面临一些新的挑战,科研环境也发生了很大的变化:首先是科学研究的问题空前复杂化,科学研究的对象已经不再是一个简单的孤立系统,比如以前化学就是化学,物理就是物理,而现在更多的跨学科的、覆盖范围更大的科研问题;另外科研过程中信息和数据的及时获取和处理显得越来越重要,仿真和大规模的计算也成为科学研究过程中分析、发现和预测的主要手段之一;这个时代也更加强调科学家之间更加密切的合作和交流,这种合作和交流可能跨学科、跨行业,也可能跨地域、跨国家。

在此形势下,为了进一步提高科学研究的效率,抢占未来科学研究的制高点,英国政府战略性地开展了以e-Science核心的虚拟科研环境建设。这种基于高速网络、海量存储、超级计算力的科研环境,使得英国各个领域的科研人员可以在具有以下特征的科研模式下开展科学研究:

1. 基于标准、面向框架。科学家及虚拟组织可以在遵循诸如WSRF(Web Service Resource Framework)、WSRP(Web Service Resource Protocol)、WSDL(Web Service Description Language)等这样一些开放标准的情况下,利用分布式的、可提供注释的工具开展科研;

2. 集成性增强。在此环境下,科研人员在资源的创建、发布、共享与发现等方面均能集成性地实现;

3. 科研资源更易于管理的。其所涉及到的存取、使用权限、归档操作均能够在安全可靠的环境下进行,从而确保科研资源的可使用性与可存取性,有利于充分地满足用户需求;

4. 个性化服务。面对不同学科科研人员和同一学科科研人员不同阶段的服务需

求，这种全新的科研环境可以提供更为个性化的服务，可由科研人员进行定制服务；

5. 智能代理。对于日常的资源需求，可通过智能代理为科研人员的日常事务提供代理，节省时间；

6. 便于操作的可扩展性。容易添加各种升级或者是用于增强功能的工具和应用插件，服务的可扩展性极高；

7. 交流的方便性。科研人员之间通过可共享交互、网格存取等沟通方式密切了彼此的交流；

8. 资源随处可得，科研随时可行。轻便的、随处可得的存取接口，利用无线的、移动的存取入口装置，科研人员可以随时随地的获取到这一虚拟科研环境中的授权使用资源。

(2) 经济利益的驱动是 e-Science 备受英国重视的直接原因

美国《福布斯》杂志的科技版“**Forbes ASAP**”曾经发表了一组文章，预言了网络时代的来临。文章指出，信息技术的下一大浪潮将在 2004—2005 年度出现，并造就 2005—2020 年 15 年的黄金时代，到 2020 年，由此产生的互联网将成长为一个 20 万亿美元产值的大产业，这一波浪潮的本质特征，就是万维网（World Wide Web，WWW）升华为网格（Great Global Grid，GGG）。

早前因为未能及时意识到 WWW 发展契机而使自身失去大好商机的英国，面对 grid 具有的强大商业潜力，这次再不重犯当年 WWW 的错误，而不是将网格视为第三代互联网加以重视（三代互联网的特征比较详见表 2-1）。于是，为了充分挖掘网格功能，把握网格带来的巨大商机，英国率先在科技领域开展试验并大力进行建设，形成了 e-Science 在英国全面发展的局势。

表 2-1 三代互联网

	年 代	名 称	特 征
第一代	20 世纪 70-80 年代	传统互联网(Internet)	计算机硬件的连接
第二代	20 世纪 90 年代	万维网(WWW)	网页的连通
第三代	21 世纪	网格 (Grid)	互联网上所有资源的全面连通

2.1.2 英国e-Science发展目标

英国 e-Science 核心计划的愿景是：让来自世界各国的科学家能够以一种前所未有的高效率工作方式，协同式地开展科学研究，并在此过程中实现数据、设备、仪器、知识等多种资源的共享。

2.1.3 工作原理

作为 e-Science 的首倡国家，英国将 e-Science 视为一种全新的科研协作模式。它建立在新一代网络技术和广域分布式高性能计算环境基础上，以互联网技术和网格

计算技术为基础，实现跨越地理界限的全球大规模数据采集、T（百万兆级）级高速计算和高性能可视化，并以此为基础将互联网的应用、高性能科学计算及资源共享提高到一个全新的层次。

2.1.4 英国e-Science建设的经费分配

2000 年 11 月，英国政府宣布投资 2.5 亿英镑用于英国的 e-Science 建设， 标志着英国大规模开展 e-Science 研究的开始。英国政府的 e-Science 研究计划主要是要解决两方面的问题：一是每个研究理事会将接受一笔专用资金，在已有基础上进行电子科学研究，解决各自领域的特定问题；二是设立跨研究理事会面临的共性问题。

根据英国政府的计划，整个英国科学和工程界的许多单位将参与执行核心计划，其中，英国的七个研究理事会均参与 e-Science 的建设计划中，同时，包括欧洲核子中心（European Organization for Nuclear Research, CERN）等在内的国际机构也参与到英国的 e-Science 建设中。

在政府提供的这笔款项中，7400 万英镑由政府科技办公室直接分配给英国的 7 个研究理事会，用于其进行各自与 e-Science 相关的研究，其中：英国经济和社会学研究理事会（ESRC）300 万；英国自然环境研究理事会（NERC）700 万；英国生物科学研究理事会（BBSRC）800 万；英国医学研究理事会（MRC）800 万；英国工程和自然科学研究理事会（EPSRC）1700 万；英国粒子物理研究理事会（PPARC）2600 万；英国中心实验室研究理事会（CLRC）500 万。1500 万用于执行跨研究理事会的 e-Science 核心计划的研究，用于开发公共研究设施。

此外，为确保研究能充分考虑工业界的利益，探索 e-Science 商业化的可能性，并为下一代电子商务建立平台，英国贸工部及英国企业界另外提供 2000 万英镑的配套资金用于 e-Science 核心计划。图 2-1 详细表明了英国 e-Science 计划项目的经费分配。

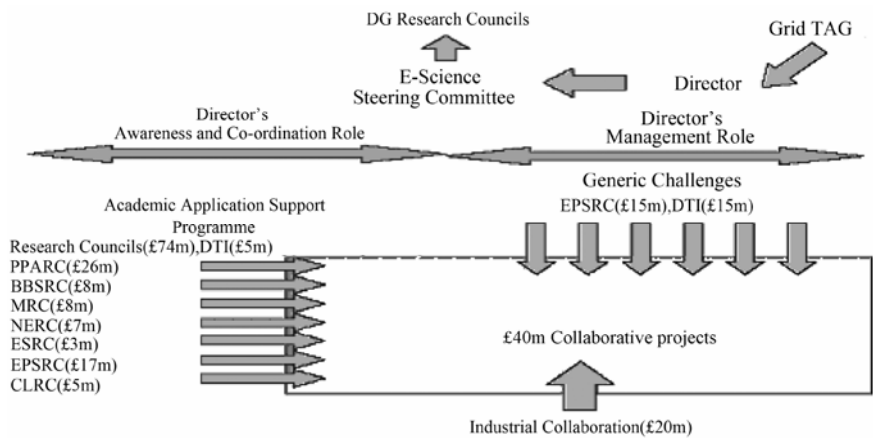


图 2-1 英国 e-Science 计划项目经费分配图

研究理事会（DG Research Councils）将确保核心计划能够与具体的应用工作紧密联系。核心计划将成为英国主要基础设施提供商——联合信息系统委员会（JISC）、英国教育研究网络协会（UKERNA）及其欧洲伙伴的主要联接点。这将保证英国科研人员可以通过英国“超级联合学术网”（Super Janet）、欧洲的 Geant 网和世界其他网络交流有足够的带宽，有利于 e-Science 的进一步开展与推广。

2.2 英国e-Science承担主体发展目标分析

2.2.1 七大研究理事会的e-Science发展目标

为了保证充分地利用 e-Science 建设奖金，英国七大研究理事会纷纷推出了自身的 e-Science 发展目标。

1. 英国科研委员会中心实验室（CLRC）的e-Science目标

CLRC（the Council for the Central Laboratory of the Research Councils）负责管理英国的大型科研设备。这些大仪器产生的数据日益成为阻碍 CLRC 发展的要素。事实上，CLRC 已经在 e-Science 方面进行一系列实验性项目。如，为欧盟网格项目建立 e-Science 实验床，将 CLRC 的中央数据存储设施升级到 Peta byte 级等等。为进一步研究 e-Science 以便其发展成为 CLRC 下一代的主要科研设施。CLRC 于 2000 年 6 月建立了一个 e-Science 中心。CLRC 将充分利用政府提供的 e-Science 资金，使该中心的实验设施进一步网格化。

此外，考虑到网格软件一般都相对较新，且尚未成熟，大多数软件来自美国，而使用者又多数是科研人员而不是软件专家。受英国研究理事会的委托，由 CLRC 牵头，爱丁堡大学和曼彻斯特大学参与，于 2001 年 7 月成立了英国网格支持中心，其作用是为研究理事会网格应用实验床项目和区域 e-Science 中心提供信息、技术支持和技术联络等，加快其使用网络软件，以确保英国的网格技术有效和快速发展。该中心是 e-Science 核心计划的重要组成部分。此外，政府用于购买新的太赫（万亿赫兹）级的计算机系统的 900 万英镑经费也由 CLRC 负责管理。

2. 英国粒子物理研究理事会（PPARC）的e-Science目标

PPARC(Particle Physics and Astronomy Research Council)计划在 3 年内投入 2600 万英镑用于 e-Science 建设，以便支持信息技术的开发。PPARC 鼓励同计算机专家及其他研究人员的合作，以满足 e-Science 的发展要求。

PPARC 于 2001 年 5 月召开了 PPARC 网格指导委员会第一次会议，确定了 PPARC 的 e-Science 项目的主要任务是开发与 PPARC 科研相关的网格和 IT 技术。同

年 7 月, PPARC 决定启动英国粒子物理网格项目(GRIDPP)和天文网格项目(AstroGrid)。对 GridPP 项目, PPARC 网络指导委员会决定第一年给予 250 万英镑的资助, 并决定为该项目预留 1700 万英镑; 对 AstroGrid 项目, 该委员会决定 3 年资助 500 万英镑。

PPARC 每年还将提供 10 个名额的 e-Science 奖学金, 从 2001 年 10 月开始, 奖金主要提供给与 PPARC 研究内容有关的从事网络计算研究的博士生。

3. 英国生物科学研究理事会(BBSRC)的e-Science目标

生物信息学是利用计算和数学方法来获取、保存、分析及解释生物信息, 以确定生物功能和工作机制及其应用。生物科学是目前世界范围内的重点研究领域。

BBSRC (Biotechnology and Biological Sciences Research Council) 计划利用政府 e-Science 资金, 加强生物信息学及网格技术在改进生活质量, 尤其是基因群组、生物结构、细胞动力学和生物多样性等方面的应用。

BBSRC 实施的生物信息和 e-Science 计划的具体目的是: 通过鼓励生物科学、生物信息学、IT 计算机科学、数学统计学、物理学和其他相关学科的相互合作, 支持开发和利用算法学软件分析方法来解决已知的生物学问题; 鼓励开发新的运算设计、软件和分析方法, 用以综合分析为一个特定目的、从不同渠道收集到的数据; 推广网格技术和软件; 支持科技界建立共同的数据标准, 使所产生的数据能被尽可能广泛地共享; 加强生物信息学和 e-Science 的培训。

4. 英国医学研究理事会(MRC)的e-Science目标

MRC (Medical Research Council) 将利用政府的 e-Science 经费, 用于生物信息学、卫生信息学和 e-Science 应用研究。MRC 的 e-Science 研究将围绕计算和数据网格的应用和中间件的研制。MRC 优先支持的领域为癌症和脑科学, 希望通过跨学科的联合(如生命科学、物理学、数学、工程学及信息技术), 共同开发和使用先进的信息技术来解决与癌症和脑科学有关的问题。

5. 英国自然环境研究理事会(NERC)的e-Science目标

NERC (Natural Environment Research Council) 的研究领域集中在地球科学。它已有 10 年开发用于改进和了解生物圈及其之间复杂的相互作用以及预测和解决环境问题的地球系统模型的经验。使用网格技术提供的崭新计算技术, 加强科学和信息技术的合作是加速该进程的重要手段。e-Science 因其有可能解决现在研究中遇到的问题而成为实现该目标的重要和及时的方法。NERC 获得 700 万英镑的资助, 用于 e-Science 的研究。2002 年 3 月, NERC 讨论了其 e-Science 支持领域, 确定了 NERC 的 e-Science 项目框架: NERC 的 e-Science 项目应尽可能与其他研究理事会、政府和

工业界合作；NERC 的 e-Science 项目不提供购买高性能计算机，EPSRC 已经准备拨款 900 万英镑用于为整个英国的 e-Science 项目购买高性能计算机；NERC 的 e-Science 项目不提供购买连接网格的基础设施；NERC 的 e-Science 项目必须能解决问题，最好两年内应有成果；有关生物信息学和太阳陆地物理学方面的 e-Science 项目包含在其他项目内，NERC 环境基因组项目有 100 万英镑用于支持生物信息学方面的研究。

6. 英国经济与社会研究委员会（ESRC）的e-Science目标

ESRC（Economic and Social Research）获得英国政府 300 万英镑的 e-Science 发展经费。主要用于开发和完善英国社会科学战略数据库。截止于 2003 年秋季，ESRC 已经启动了十一个项目，并于同年成立了英国国家社会科学 e-Science 研究中心（NCeSS），该中心将自身的研究任务划分为两个阶段。第一阶段侧重于建立一个英国境内社会科学 e-Science 研究协调中心；第二阶段突出强调传播英国社会科学 e-Science 的研究成果。

7. 英国工程和自然科学研究理事会（EPSRC）的e-Science目标

EPSRC（Engineering and Physical Sciences Research Council）在英国的 e-Science 研究中起着十分重要的作用，它总共得到政府 3 笔资助。其中 1700 万英镑用于实验床在工程和自然科学中的应用，主要用于计算和数据网格应用及中间件的开发；900 万英镑用于购买高性能计算机；1500 万英镑由 EPSRC 代表各研究理事会用于开发 e-Science 通用核心技术。此外，DTI（Department of Trade and Industry）和英国工业界各支持 EPSRC2000 万英镑，用于帮助 e-Science 核心技术向商业化方向转换。EPSRC 目前已经完成网格实验床项目的征集，正在进行项目评审工作。EPSRC 打算支持的项目很少（6 到 8 个），所有项目必须保证 3 年内开发出一个网格工作系统，必须基于开放源软件（如 Globus），必须有工业界的支持。

2.2.2 国家e-Science研究中心的e-Science发展目标

为了引导英国 e-Science 的发展路线。具体包括：领导、合作、国际参与、培训、激励、建立共同体、研究与发展，在英国研究总理事会的支持下，英国在爱丁堡设立一个国家 e-Science 中心，该中心热衷于合作项目的讨论，并把自身所面临的挑战定义为，通过快速的传递理解以及对教育和实践的支持，给英国的 e-Science、工程和医学做出贡献。

英国国家 e-Science 中心的发展目标就是，激励与维持英国 e-Science 的发展，更为重要的是，致力于它的国际化发展以及保证它的技术得以迅速地向工商业传播；在苏格兰科学家的努力下，鉴别与支持苏格兰境内各研究所之间及研究所内部的数字科研项目，同时提供合适的技术基础设施和支持，以保证数字科研技术得到迅速提升；鼓励计算科学研究与 e-Science 应用之间的相互交流及思想的双向流动；促进

科学资料法治方法与分析上的提高，并成为管理、共享及对研究；数据进行最好利用的、质量最高的系统与数据仓库的主要来源。

2.2.3 十大地区中心的e-Science发展目标

1. 南安普敦地区e-Science中心（Southampton e-Science Centre, SeSC）

南安普敦地区中心的目标是“发展成为南安普敦地区的一个优秀的 e-Science 中心，努力成为更大范围社区的人们提供一个共享基础设施和咨询经验的中心。”为了落实上述目标，该中心确定了自身的 e-Science 发展任务。即确保南安普敦地区能够得到 GRID 技术的支持，而且准备参与到国家或国际方面的网格活动，如 Globe Grid Forum（全球网格论坛）。尽量提供那些可以共享的基础设施如 e-Science 项目所需的，Access Grid 及对社区发展有利的其他网格技术设施。配置项目管理来支持 e-Science 活动。从研究委员会、DTI 及其他的实业资源中争取进一步的赞助。

2. 牛津地区e-Science中心（Oxford e-Science Centre, OeSC）

牛津地区 e-Science 中的目标是支持当地优秀地区中心的发展。为此，它将通过 SuperJanet 连接国家网格，并提供主要的超级计算机设施，根据 Globus 标准使用相应的网格软件包。

3. 西北地区e-Science中心（e-Science North West, eSNW）

西北地区 e-Science 中心的目标是：赞助西北地区的一些项目，这是为了确保能够快速实现网格资源的共享；给西北地区那些正在实施的一些项目提供适当的技术设施支持，以便让西北地区能够成为知晓网格、使用网格的中心（Grid Awareness and Grid Enabled）；寻找合作伙伴，以便能够一起研究基本的网格技术并通过链接把网格技术应用到 e-Science 中，学术家与实践家一起合作也可以为项目找到一定的支持者，科学家与情报学家一起分享共同的利益；在生物医学的 e-Science 领域开发一种特定的资源，并且努力成为一个高质量知识库的原始来源。

为了落实自身上述目标，该中心确定了自身的 e-Science 发展任务。即，与工业伙伴共同合作赞助部分 e-Science 项目，同时负责研发一些应用于西北地区的网格技术；成为西北地区网格和国家二级网格之间的一个结点。

4. 东北地区e-Science中心（North East Regional e-Science Centre, NEReSC）

东北地区 e-Science 中心为高校和工业界提供 e-Science 的专门技术，倡导一批企业化的项目，并力争在以下方面取得发展：成为 e-Science 领域的优秀研究中心；发起和管理一百万英镑的关注工业界匹配建设的项目；在 e-Science 地区中心中充当公司和高校研究团体的联系纽带；开展 e-Science 相关领域的培训、交流活动；参与计算/数据资源和设施的国家网格计划。

为了落实自身上述目标，该中心确定了自身的 e-Science 发展任务。即，引领工业界和学术届伙伴确定、资助和支持高性能的项目。设计基于网格的系统，教育和培训。支撑、开发一般的中间件、网格系统的安装、配置和管理。

5. 贝尔法斯特地区e-Science中心（Belfast e-Science Centre, BeSC）

贝尔法斯特地区 e-Science 中心的 e-Science 发展目标就是，促进北爱尔兰地区的网格技术迅速提高。为了落实自身上述目标，该中心确定了自身的 e-Science 发展任务。即，承担和主办 e-Science 工业项目；为北爱尔兰本地公司和其他组织提供相应的技术框架和资源，支持 e-Science 的一系列的活动的；为对网格技术感兴趣的本地高校，公司和其他组织提供专家建议和咨询。

6. 剑桥地区e-Science中心（Cambridge e-Science Centre, CeSC）

剑桥地区 e-Science 中心的目标定位是，持剑桥地区的包括科学家和产业在内的 e-Science 项目；通过使用网格（Grid-enabled）来解决 Terascale 问题，推动新的科学发展；开发新型基于网格的工具，用来解决大型数据处理，在更广泛区域网络中的高性能计算和可视化应用。

剑桥 e-Science 中心由英国贸易工业部为了促进东英格兰 e-Science 研究而批准成立，它是由 SuperJanet 连接的国家网络的一部分，通过使用基于 Globus 标准的网格软件包，为 e-Science 研究人员提供获取大量具有超级计算力的设备的使用权限。

7. 英国研究理事会中心实验室理事会e-Science中心（CLRC e-Science Centre, CLRCeSC）

英国研究理事会中心实验室理事会 e-Science 中心的目标是给英国主要的科学设备及其使用群体开发了一种基于网格的基础设施。它将负责对 CCLRC 网格发展计划进行管理，并协调英国数字科研网格的发展。

8. 伦敦地区e-Science中心（London e-Science Centre, LeSC）

伦敦地区 e-Science 中心的目标定位是，建立和维护基于 e-Science 控制的网格启动装置的网格环境；设置工业合作项目的程序，用来发展和配置一般网格技术；促进网格和 e-Science 技术的知识；在网格和 e-Science 方面作为本地建议和培训资源；在国家网格上贡献基础构架。

为了实现上述目标，该中心确定了 e-Science 发展任务。即通过与数学模型，量子物理，生物信息，环境模型和工程等的科学家的合作，支持伦敦和东南地区的 e-Science 活动的顺利开展。

9. 威尔士e-Science中心（Welsh e-Science Center）

威尔士 e-Science 中心的目标定位是,在威尔士和英国西南地区提高 e-Science 研究和发展;鼓励行业 and 高校开发 e-Science ICT 技术(如网格计算)跟踪研究数据和知识管理、异类分布信息资源的互操作、科学和工程应用解决问题的环境以及虚拟和合作应用的分布式可视化等领域的热点问题。

10. 英国网格支持中心Grid Support Center （Grid Support Centre, GSC）

英国网格支持中心作为英国贸工部所发起的 e-Science 核心计划的一个部分,它对网格中间件的配置、运作和维护各个方面提供支持,更给英国网格测验床分布式资源的管理所涉及到的各个方面提供帮助。在它这一中心的网站页面提供了各种与英国建立计算网格有关的文件、软件、网页链接的汇集。它还开发各种网格启动工具（Grid Starter Kit）。

2.2.4 七大e-Science优秀研究中心

1. 莱斯特e-Science中心（The Leicester e-Science Center）

该中心成立于 2003 年 9 月,中心在天体物理学与宇宙方面具有独特的理论优势,但是它也必须适应更广泛范围内的 e-Science 活动的开展。中心必须为地方的学术共同体提供 e-Science 有关的专家技能,使得研究人员可以充分利用因特网的新成果,参与合作的企业可以访问到数据集、大型计算资源和高性能可视化组件。

莱斯特 e-Science 中心,也是英国目前七个优秀的 e-Science 中心之一,它的成立是为了对英国当前的 e-Science 国家和地区研究中心形成强有力的补充,并给 e-Science 的应用和英国 e-Science 网格的未来资源提供额外的专家技能。与英国其他 e-Science 中心一样,它也得作为支持远程会议的访问网格（AccessGrid）节点的一个主办者之一。该访问网格使得人们可以通过因特网获得高质量的音频与视频通信。这将促进来自多个领域的人们进行合作,而且也必将可以将这种通讯广泛应用到会议、培训、研讨会和讲座中去。

2. UCLGrid/HPC

UCL Grid/HPC 作为提供给 e-Science 研究人员交流使用的论坛,它将 e-Science 研究者与对计算网格的技术和应用感兴趣的服务提供商汇集到一起。它的目标就是让网格服务的配置与发展在高校中顺利进行并为网格服务的配置与发展方面的合作提供便利,并凭借伦敦大学学院（University College London, UCL）的研究优势,开拓各学科之间的协调发展。

3. The White Rose Grid

The White Rose Grid 将自身目标定位为提供与传递网格技术与服务的国际领导者。为此，它将自身的发展任务细化为五个方面：

① e-Science 研究：在 e-Science 的研究上加强与企业的合作关系。在 e-Science 的研究上，重点关注的核心领域是：实时的分布式设计支持、诊断服务、实时的分布式合作可视化服务；

② 科学共同体：支持和扩大新的科学共同体，包括生物技术、航空宇宙、组织工程（tissue engineering）和卫生保健；

③ 追求超越：通过建立起可以让区域高校可以获取的并与基于公司而建立起来的科学相适应的区域网格设施，发挥 White Rose 在研究上的优势，同时弘扬与这种优秀之处密切相关的知识；

④ 市场评估：WRG（White Rose Grid）将与约克郡促进会（Yorkshire Forward）和 IT 合作伙伴开展合作，忠实地对网格技术的潜在区域需求进行评估；

⑤ 网格与其他共同体：中心的目标就是成为世界上的高校网络与其他国家（如英国 e-Science）、国际网络中一名富有成效的合作者。

4. 里丁大学e-Science中心

作为英国当前七大卓越的 e-Science 中心之一，里丁大学 e-Science 中心将自身的目标定位为，通过与学术界、政府机构（如 Met Office，即英国气象局）和企业界的合作，开展示范项目，以便在环境科学共同体内推广、宣扬 e-Science 的方法。里丁大学 e-Science 中心希望，通过这些示范项目来表明，e-Science 的巨大潜力就是成为一种相当有用的方法。这种方法可以被许多领域的环境科学家在日常的科学研究活动中加以使用。同时，通过对这种方法的使用，那些在以往看来是难以从事、甚至是费时的作业，如对极大规模数据集的共享与操作，人们将可以更高效率地执行它。

里丁大学 e-Science 中心相信，通过开展 e-Science，获益的并不仅仅是科学研究共同体。对于许多公共和私有部门来说，他们的工作之所以能够正常地开展，很大程度上依赖于环境数据。许多情况下，倘若这些公共和私有部门能够持续不断地获得实时、最新的环境数据，如天气模式和洋流的预测数据，他们的工作效率将明显得到改善和提高。

里丁大学 e-Science 中心就是通过对这些共同体（包括公共和私有部门、科学研究共同体）的深入接触，鉴别他们对于这方面数据的需求，并向他们提供建议和解决方案。

5. 兰开斯特e-Science中心（Lancaster e-Science Center）

针对社会科学量化问题的前沿性研究要求将各种方法加以综合，才能构建处理复杂的社会与经济问题的模型。一个综合的处理模型意味着多种相互领带关系存在于运行过程中。e-Science 将有利于促进基于实证理论的证据创新，使得决策者人员能够以最为影响力的政策手段做出更为科学合理的决策从而提高政策效力。兰开斯特 e-Science 中心的发展目标就是，加快开发出面向社会科学的、新型计算与数据管理基础设施，为多个学科领域的国内合作以至国际合作提供支持。为此，该中心所开展的一系列活动，重点在于促进来自管理、社会科学以及其他相关学科的研究人员可以建立起一种模型，并使得该模型可以揭示出社会与经济现象的全面复杂性。

6. 中部e-Science中心（Midland e-Science Center）

中部 e-Science 中心是建立在英国伯明翰大学计算机学院的一个虚拟中心。该中心长期关注与网格相关的高效率算法和软件技术的开发（包括仿真、数值计算、进化计算方法），并将其应用到能够促进科学发展和让业界获益的复杂问题上，成果卓著。

中部 e-Science 中心的工作重点主要集中在四个方面：给中部地区提供网格可获取性与连通性；给极大型复杂系统的计算模型化活动创造研究的聚焦点并工业的挖掘相关专家技能的源泉；对数值计算和网格仿真技术进行长期深入研究；鼓励不同科学、工程的不同学科与业界之间开展合作。

7. 布里斯托尔e-Science中心（Bristol e-Science Center）

布里斯托尔 e-Science 中心致力于在物理、医学、天文学、地球科学、生物科学等多个自然科学领域推广网格技术。为此，自成立以来，该中心已经在 BioSimGrid 项目、CPOM 项目、eSTAR 项目、GENIE 项目、GridPP 项目、MammoGrid 项目、ROSE 项目、VidGrid 项目等大小型 e-Science 项目中发挥了重要作用。

2.3 英国e-Science技术框架与技术模块

2.3.1 英国e-Science组织框架

2003 年 5 月，来自英国的 D. Berry 等学者首次提出，作为虚拟研究环境的代名词，要实现 e-Science 的顺利实施与推广，必须构建发展 e-Science 必不可少的组件架构。在此基础上，他们首次提出了图 2-2 所示的 e-Science 的组织框架。



图 2-2 e-Science 组织框架

2.3.2 三十一个技术组件

目前，完善 e-Science 技术基础架构是世界各国家或地区现阶段 e-Science 的建设聚焦点。在 e-Science 建设方面走在世界前列的英国更是充分地认识到技术基础架构的重要性。在实施 e-Science 核心计划的过程中，英国国家 e-Science 中心确定了发展 e-Science 所必需的三十一个技术组件，并辅以九十一个试验项目展开可行性研究。它所确定的发展 e-Science 所需的三十一类技术组件包括：应用服务；网格计算服务；领域本体、任务本体与元数据；发布与归档服务；问题求解环境；合作工具包；数据管理服务；数据挖掘与分析服务；网格管理服务；仪器管理服务；市场经济服务；模式与程序原型；时序安排、监测与诊断服务；虚拟与沉浸环境服务；可视化服务；工作流服务；审核；认证与授权；代理；合作与远程仪器服务；协同调度；数据访问服务；数据编目与溯源；容错管理；全局事件服务；网格信息服务；全局请求；监测；服务质量的可靠性；资源访问服务；安全与隐私。

2.3.3 四层架构的建设模式

由英国工程与自然科学研究理事会（Engineering and Physical Sciences Research Council，EPSRC）投入三百多万英镑于 2002 年启动的 DAME（Distributed Aircraft

Maintenance Environment）项目将研究目的定位为，给分布诊断建造一个网络试验，而它的应用示范品将是一个由劳斯莱斯（Rolls Royce）及其信息系统伙伴——数据系统及解决方案公司的需求所推动的、分布式航行器维护环境。

在该项目中，科研过程被构思为由图 2-3 所示的四层架构。在该模式中，最底层为网络基础设施，负责调用网格中各种可用的资源；处于网络基础设施之上的是网络中间件层，它主要负责网络服务的管理，需要提供包括资源代理、数据代理、任务调度以及服务注册在内的各种服务，是 DAME 系统层与网络基础设施“打交道”的媒介。第三层的是 DAME 系统层，它通过利用各种网格工具软件与网络中间件进行联系，并建立 DAME 诊断门户系统将工作流支持服务、决策支持服务以及资源支持服务以集成的方式提供给用户使用。在第四层，也即是处于该模式最顶层的“维护空间”，这是用户进行维护操作以及开展商务活动的操作层。在这一层中，用户通过诊断门户系统对各种用于维护的个案数据进行操作。

在 DAME 的科研模式中，其虚拟组织也表现出鲜明的特征，具有汇集了高性能的资源、资源性质广泛多样、实施本地化管理、访问可控制、安全政策保障完善等被 e-Science 所推崇的优点。

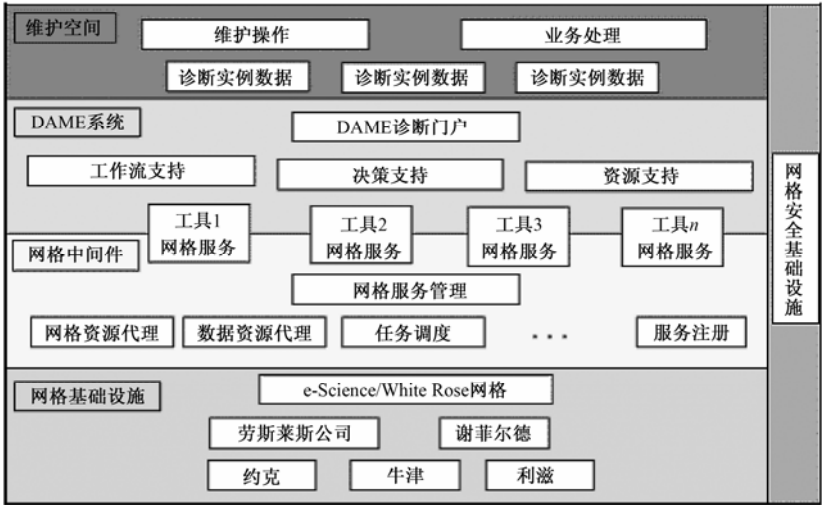


图 2-3 基于网格的 e-Science 全新科研协作模式（以 DAME 为例）

2.3.4 二十六项技术模块

作为一种全新的科研协作模式，e-Science 的顺利实现，需要有多种技术的支撑，从英国所构造的 e-Science 组织框架到其在建设过程中提出的三十一类技术组件以及四层架构的建设模式，无不充分说明这一点。国内学者在介绍英国 e-Science 发展状况时指出，e-Science 的技术范围相当广泛，其中，有二十六种技术是 e-Science 发展

所必不可少的。它们是：资源发现，即提供资源名称及特点，以便在分布式系统中对资源进行自动定位；同步和协作，即如何协作利用各种资源和进行复杂协调计算；容错性，即如何在分布式系统中出现各种意想不到的错误情况下保证任务的正确执行；安全性，即确保在 Internet 环境下和大量用户的情况下保证信息及资源的安全；并发性和一致性，即保证分布式异构系统的数据的一致及操作的原语性；性能，即在分布式大规模的系统中保持适当的性能；异构性，即分布式系统的异构，包括网络、硬件、软件、功能、协议等；可扩展性，即扩展到成千上万个节点后，仍能正常工作；数字版权，即科学内容的数字版权管理；溯源创建和管理，即为了便利地重复使用信息及重复实验记录、创建和管理数据产生的时间；元数据及其描述工具；服务描述及工具；工作流描述及工具；新设备的信息获取和表示；发布和定位知识水平的服务语言；提供大规模本体服务的方法与工具；注解服务；知识获取工具；从大量内容中动态搜索需要的内容；自然语言处理；推理服务，知识发现服务；多种知识服务的协作；在无所不在的设备中嵌入知识服务；面向服务的体系结构；基于 Agent 的技术；网络协议研究；e-Anything 和 e-Science 的关系，等等。

2.4 英国e-Science当前进展

2.4.1 英国e-Science核心计划

2001 年开始的英国 e-Science 计划是近年来英国政府在信息技术领域支持的最大的研究计划，总经费接近 2.5 亿英镑。英国政府把它作为改变英国科学研究现状的重大举措。

e-Science 意味着通过全球分布协作，使用海量数据集合、万亿次级的计算资源和高性能的可视化提高科学研究效率。2002 年，英国首相布莱尔曾指出，“英国是第一个开发全国范围的 e-Science 网络的国家。这样的网络使得我们能像用 web 访问信息一样容易地访问计算能力、科学数据仓库和实验设施。”英国 e-Science 倡导者 John Taylor 谈到 e-Science 的访问分布存储的意义时指出：e-Science 意味着通过全球分布协作，使用海量数据集合、万亿次级的计算资源和高性能的可视化，有效地提高科学研究效率。

为了确保 e-Science 的顺利开展，在科技办公室（OST）和英国贸工部（DTI）的大力支持下，英国于 2001 年开始实施 e-Science 核心计划，主要用于支持基础中间件开发、e-Science 中心和英国 e-Science 网络建设，促进技术辐射和国际合作。自实施以来，e-Science 核心计划已经吸纳了 80 个以上的企业参与研究，包括 IBM、Sun、Intel、微软，终端用户的 Rolls Royce 和西门子，以及 NAG, Cybula 这样的中小企业，他们共计投入资金超过 3000 万英镑。

英国 e-Science 核心计划共分两个阶段：

第一阶段的工作重点是：提出网络的研究和开发计划，支持 e-Science 实验床。持

续时间为 2001 年到 2004 年，经费 1.2 亿英镑。其中 7500 万支持各个领域的网格应用，3500 万支持核心研究计划，另有 1000 万英镑用于高性能计算机等重要装备采购。网格应用项目作为先导项目，分布在粒子物理和天文学、生物技术和生物学、医学、自然环境、经济和社会学、工程和物理科学等领域。核心计划第一阶段建设了国家 e-Science 中心（NeSc）以及伦敦、剑桥、牛津、南安普敦等 13 个 e-Science 地区中心和伯明翰/瓦威克建模中心、布里斯特尔媒体技术中心等 7 个优秀的 e-Science 研究中心。

第二阶段的工作重点是制订长期研究目标，突出网格中间件、可持续运行的网格服务环境、网格支持环境中心、数据维护中心等方面的研究和建设。持续时间为 2003 年到 2006 年。从第二阶段开始，e-Science 核心研究计划将突出网格中间件和 e-Science 网格的建设。为了加强运行管理，将成立一个网格支持中心，对网络提供统一服务。另外，成立开放中间件基础设施研究所，专门从事网格中间件的研究、开发和工程化。到 2005 年底，第二阶段已经在南安普敦大学建设了开放中间件基础设施研究所，在 RAL（Rutherford Appleton Laboratory，阿普尔顿实验室）国家实验室建设了网格支持中心，建成了由牛津、RAL 等六个中心组成的英国国家网格服务环境，在爱丁堡大学建设了数据维护中心。同时启动支持了一批新的网格应用项目。

英国 e-Science 计划计划在 2006 年结束。从 2006 年起，英国将把 e-Science 作为一种常规内容列入各个研究理事会的科研计划，扩大在各个科学领域的应用。工程和物理科学研究理事会仍将继续主持 e-Science 核心计划，为所有学科提供技术和平台支持。在英国，e-Science 已经为大家所认可，对它进行常规的、日常的支持，而不再以一种特别手段来强调它的特殊性或重要性，这表明英国的 e-Science 建设正逐步走向成熟。通过实施这一核心计划，英国整个国家在研究水平、科技竞争力和将来的研究模式方面都取得了重大的进步。

目前，英国 e-Science 建设项目不断涌现。不仅国家 e-Science 中心身先士卒，大力启动 e-Science 项目，包括 7 个优秀研究中心 7 个、10 个地区 e-Science 研究中心 10 个等在内的多家研究机构均纷纷加入到 e-Science 的项目建设中。

2.4.2 英国国家e-Science中心参与的e-Science项目

英国国家 e-Science 中心是英国 e-Science 研究的国家中心。在该中心的网站上，中心将其所参与的 e-Science 项目从组件、项目成果的潜在应用领域、项目的类型（根据项目运作模式划分为五种类型）、经费来源机构进行了聚类。通过这些聚类，全世界对 e-Science 感兴趣的人员均可以对英国国家 e-Science 中心在项目方面的发展情况了如指掌。

1. 从组件角度聚类e-Science项目

英国国家 e-Science 中心从组件角度对 e-Science 项目进行聚类。它根据项目涉及

到的组件分为三层。第三层包括五个小类，第二层包括十一个小类，而第一层包括的小类总数在三层中居首，共计有十五个小类。各小类都有数目不一的项目聚集。具体分布情况见表 2-2。

从表 2-2 中可见，目前，英国国家 e-Science 中心所从事的 e-Science 项目在组件开发方面，主要集中在以下七个方面：应用服务；数据管理服务；数据挖掘与分析服务；合作工具包；网格计算服务；领域本体、任务本体与元数据；可视化服务。从数量上分析，在没有进行去重之前，组件这一聚集角度所聚集的项目总数达到 332 个（截止于 2005 年 8 月 27 日）。而经去重后，此角度下聚集的项目数仅有 91 个。

表 2-2 从组件角度聚类 e-Science 项目

所 属 层 次	小 类 名 称	项目总数（未去重）
第三层（99）	应用服务	45
	网格计算服务	20
	领域本体、任务本体与元数据	19
	发布与归档服务	1
	问题求解环境	14
第二层（147）	合作工具包	23
	数据管理服务	34
	数据挖掘与分析服务	29
	网格管理服务	12
	仪器管理服务	7
	市场经济服务	4
	模式与程序原型	1
	时序安排、监测与诊断服务	11
	虚拟与沉浸环境服务	4
	可视化服务	17
	工作流服务	5
第一层（86）	审核	3
	证书与授权	6
	代理	6
	合作与远程仪器服务	7
	协同调度	8
	数据访问服务	9
	数据编目与溯源	8
	默认管理	4
	全局事件服务	5
	网格信息服务	4
	全局请求	7
	监测	4
	服务质量的可靠性	7
	资源访问服务	2
	安全与隐私	6

2. 从项目成果的潜在应用领域聚类e-Science项目

英国国家 e-Science 中心从项目成果潜在的应用领域对 e-Science 项目进行聚类。到目前为止，它共分十大领域进行聚类，各领域下面均有数量不一的 e-Science 项目。具体分布情况见表 2-3。

表 2-3 从项目成果的潜在应用领域聚类 e-Science 项目

学 科 领 域	各学科包含的项目总数（未去重）
生物科学	35
商业领域	0
计算科学基础	0
工程与物理学	21
环境领域	9
基础架构	2
医学	21
中间件	38
粒子物理与天文学	11
社会学领域	1

从表 2-3 中可见，目前，英国国家 e-Science 中心所从事的 e-Science 项目的成果的潜在应用领域主要集中在生物科学、工程与物理学、医学和中间件四个方面。从数量上分析，在没有进行去重之前，项目成果潜在的应用领域这一聚集角度所聚集的项目总数达到 138 个（截止于 2005 年 8 月 27 日）。而经去重后，此角度下聚集的项目数仅有 115 个。

3. 从项目类型聚类e-Science项目

英国国家 e-Science 中心从项目类型的角度对 e-Science 项目进行聚类。它根据项目运作模式将 e-Science 项目划分为七种类型。这七种类型即是：核心工业项目（centre industrial projects）、试点项目（pilot projects）、跨学科研究合作项目（IRC projects）、示范项目（demonstrator projects）、国际合作项目（international projects）、公开征询项目（open call projects）和综合性项目（Miscellaneous e-Science Projects）。这些类型各自所汇集的项目数量如表 2-4 所示。

表 2-4 从项目类型聚类 e-Science 项目

项 目 类 型	各类型包含的项目总数（未去重）
核心工业项目（centre industrial projects）	49
试点项目（pilot projects）	59
跨学科研究合作项目（IRC projects）	6
示范项目（demonstrator projects）	15
国际合作项目（international projects）	10
公开征询项目（open call projects）	20
综合性项目（Miscellaneous e-Science Projects）	73

从表 2-4 中可见，目前，在跨学科研究合作项目方面，英国国家 e-Science 中心所从事的 e-Science 项目在这方面比较薄弱。尽管 e-Science 强调全球领域内的科研协作，但是，到目前为止，就 e-Science 项目本身而言，在开展国际合作方面，即便是对 e-Science 非常热衷的英国，在这方面也未有充分地反映。理由是：从项目类型这一角度统计英国国家 e-Science 中心从事的 e-Science 项目总数，英国国家 e-Science 中心从事的国际合作项目仅仅占它所从事的各类型项目总数的 4.3%。不过从表中可知，英国国家 e-Science 中心目前所启动的 e-Science 项目依然是以试点为主，而示范性项目则为数不多。从数量上分析，在没有进行去重之前，项目成果潜在的应用领域这一聚集角度所聚集的项目总数达到 232 个（截止于 2005 年 8 月 27 日）。而经去重后，此角度下聚集的项目数是 232 个。

4. 从经费来源机构聚类e-Science项目

英国国家 e-Science 中心从项目的经费来源机构这一角度对 e-Science 项目进行聚类。它根据当前项目的十八个经费来源机构，包括：AHRB（Arts and Humanities Research Board）、BBSRC（Biotechnology and Biological Sciences Research Council）、Core（e-Science Core Program）、DOE（Department of Energy）、DTI（Department of Trade and Industry）、EPSRC（The Engineering and Physical Sciences Research Council）、ESRC（The Economic and Social Research Council）、EU（European Union）、HEFCE（The Higher Education Funding Council for England）、HEFCW（Higher Education Funding Council for Wales）、JISC（the Joint Information Systems Committee）、MRC（Medical Research Council）、NERC（Natural Environment Research Council）、NIH（National Institutes of Health）、PPARC（Particle Physics and Astronomy Research Council）、SHEFC（Scottish Higher Education Funding Council）、Wellcome Trust，将项目分十七个类目进行归类。各个资助机构所资助的项目数量如表 2-5 所示。

表 2-5 从经费来源机构聚类 e-Science 项目

机 构 名 称	机构 简称	机 构 全 称	项 目
英国人文学科研究委员会	AHRB	Arts and Humanities Research Board	0
英国生物技术与生物科学研究委员会	BBSRC	Biotechnology and Biological Sciences Research Council	22
英国 e-Science 核心计划	Core	UK's Core e-Science Programme	101
英国能源部	DOE	Department of Energy	0
英国贸易工业部	DTI	Department of Trade & Industry	7
英国工程与自然科学研究理事会	EPSRC	Engineering and Physical Science Research Council: EPSRC	52
英国国家经济与社会研究理事会	ESRC	Economic and Social Research Council	10
欧盟	EU	European Committee	13
英国高等教育资助委员会	HEFCE	Higher Education Funding Council for England	1

机 构 名 称	机构 简称	机 构 全 称	项 目
威尔士高等教育基金管理委员会	HEFCW	Higher Education Funding Council for Wales	0
英国联合信息系统委员会	JISC	The Joint Information Systems Committee	6
英国医学研究理事会	MRC	Medical Research Council	9
英国国家自然环境研究委员会	NERC	Natural Environment Research Council	6
NFS	NFS	NFS	0
美国国家医学研究院	NIH	National Institutes of Health	1
英国粒子物理与天文学研究理事会	PPARC	The Particle Physics and Astronomy Research Council	19
苏格兰高等教育基金管理委员会	SHEFC	Scottish Higher Education Funding Council	1
英国 Wellcome Trust 基金会	Wellcome	The Wellcome Trust	2

从表 2-5 中可以容易地得出，英国 e-Science 核心计划在英国国家 e-Science 中心开展 e-Science 项目方面有着支柱的作用，单从经费来源机构这一角度所统计出来的 238 个项目中，英国 e-Science 核心计划资助的项目就达到 101 个，将近半数。另外，英国工程与自然科学研究理事会、英国生物技术与生物科学研究委员会和英国粒子物理与天文学研究理事会也是英国国家 e-Science 中心开展 e-Science 项目的主要投资者。正是得益于上述机构的支持与资助，英国的 e-Science 研究项目才能如雨后春笋般地涌现，并在短时间内从发展规模与影响力上独占了国际鳌头。而从机构性质上分析，教育领域的机构也开始将触角伸向 e-Science 项目的资助领域，但目前来看，其影响力毕竟甚微。诚如表 2-5 反映的情况，苏格兰高等教育基金管理委员会、英国高等教育资助委员会和威尔士高等教育基金管理委员会资助的项目数分别为 1 个、1 个和 6 个，这在英国当前规模庞大的 e-Science 项目群中所占的份额是很小的。

2.4.3 英国其他e-Science中心参与的e-Science项目

截止于 2005 年 8 月 29 日，英国在开展 e-Science 项目研究的过程中，除设立国家 e-Science 中心后，英国在国内的各个地区、大学设立了地区 e-Science 研究中心，数目达到了十个。为了发挥示范作用，鼓励先进，在此十个地区 e-Science 中心之外，英国还设立七个优秀的 e-Science 中心。此外，鉴于大型 e-Science 项目对于后续 e-Science 研究的借鉴性影响力极大，英国还特地设立了七个以项目为依托的 e-Science 中心。

1. 英国十个地区e-Science研究中心所参与的e-Science项目简介

截止于 2005 年 8 月 29 日，英国已经成立了十个 e-Science 研究地区中心。由于各中心挂靠的研究单位的研究实力表现各异，因此，它们各自所承担涉足的 e-Science 研究项目的规模也就大小不一。表 2-6 揭示了英国各个 e-Science 地区中心承担过的项目数。

表 2-6 英国各个 e-Science 地区中心承担过的项目数

地区中心名称	Center name	项 目 数
贝尔法斯特地区中心	Belfast e-Science Centre (BeSC)	13
剑桥地区中心	Cambridge e-Science Centre (CeSC)	12
国家中央实验室委员会数字科研中心	CLRC e-Science Centre (CLRCeSC)	57
西北地区中心	e-Science North West (eSNW)	11
网格支持中心	Grid Support Centre (GSC)	8
伦敦地区中心	London e-Science Centre (LeSC)	16
东北地区中心	North East Regional e-Science Centre (NEReSC)	13
牛津地区中心	Oxford e-Science Centre (OeSC)	27
南安普顿地区中心	Southampton e-Science Centre (SeSC)	22
威尔士数字科研中心	Welsh e-Science Centre (WeSC)	32

从表 2-6 中可知，仅是从项目数量的角度分析，英国国家中央实验室委员会数字科研中心所从事的 e-Science 项目数在是英国十个 e-Science 地区研究中心中的佼佼者，达到 57 个，其数目仅次于英国国家 e-Science 中心。而其他地区 e-Science 中心所从事的 e-Science 项目数虽然没有英国国家中央实验室委员会数字科研中心从事的 e-Science 项目数多，但也各有特色。经统计，上述十个地区 e-Science 中心所从事的项目数在没有去重的情况下达到了 209 个。

2. 英国七个优秀e-Science研究中心所参与的e-Science项目简介

与十个地区 e-Science 研究中心相比，英国七个优秀 e-Science 研究中心在 e-Science 项目的研究方面，各个地区 e-Science 中心更注重项目的质量建设，因此，相比之下，七个中心所从事的 e-Science 项目总数相对较少些。各个优秀 e-Science 中心承担的项目数如表 2-7 所示。

从表 2-7 中可知，尽管英国各优秀 e-Science 研究中心将项目质量建设看得很重，但在数量建设上，与其他 e-Science 地区中心相比毫不逊色者，当是伦敦大学学院数字科研中心莫属。

表 2-7 英国七个优秀 e-Science 研究中心所参与的 e-Science 项目

优秀 e-Science 中心名称	Center name	项 目 数
白玫瑰网络	The White Rose Grid	1
伦敦大学学院数字科研中心	University College London(UCL)	31
布里斯托尔大学数字科研中心	Bristol	1
兰开斯特大学中心	Lancaster	5
伯明翰大学的中部数字科研中心	Birmingham	8
里丁大学数字科研中心	READING	11
莱斯特大学数字科研中心	Leicester	1

3. 其他 7 个 e-Science 研究中心涉及的 e-Science 项目

为了便于项目成果性质相近的 e-Science 项目之间的协调、沟通，英国还设立了七个 e-Science 虚拟中心，它们是：Access Grid Support Centre (AGSC)、Digital Curation Centre (DCC)、National Grid Service (NGS)、National Centre for e-Social Science、National Centre for Text Mining(NaCTeM) National Institute for Environmental e-Science (NIEeS) 和 Open Middleware Infrastructure Institute(OMII)。由于这些中心所涉及的 e-Science 项目已经包括在前面统计过的项目中，因此，若仅是从 e-Science 项目总数的角度研究英国 e-Science 项目相关情况的话，则若再将这七个虚拟中心的项目数进行列举，显得意义不大。

2.4.4 典型 e-Science 项目进展

1. MyGrid 项目

英国曼彻斯特大学的生物信息学数据库和计算网格 MyGrid，为生物学家研究人员提供了一个基于网格的实验环境，研究者可以方便地经由互联网进入该系统，使用系统专有的果蝇和酵母数据仓库和基因组功能注释平台。该系统的研发考虑了网格的应用和进一步拓展，从数据库构架、标准、数据库接口着手，结合生物学的本体论约定，充分应用了网格中间件技术、Web 技术和网格计算技术，可以说是 e-Science 研究中生物信息技术与网格技术结合的一个相当成功的研究和应用成果。

2. Discovery Net 项目

英国帝国理工大学的基于高性能计算网格，开展生物学知识发现研究的 Discovery Net 研究项目，基于帝国理工大学计算机科学系生物信息技术研究组 8 年以上的研究成果—生物信息知识发现技术平台 KDE，获得了 e-Science 核心计划为期三年、经费为 208 万英镑的资助，并且是作为英国国家级战略性研究项目。该项目

的研究成果已经成功地运用在美国、欧洲一些大型科研机构、生物技术和医药企业中，其技术体系整合异构生物学数据库、整合各类生物信息学分析软件和算法，以实现大规模数据挖掘的能力和其兼容性、可扩展性。在根据研究单位、实验室、生物技术、医药产业的具体需求所提供生物信息学解决方案的灵活性和能力方面，该项目的成果都是世界一流的。

2.5 小 结

据不完全统计，截止于 2005 年 11 月底，英国已经完成以及正在开展的 e-Science 项目总数已经超过三百个。在如此庞大的项目群中，英国 e-Science 项目管理者始终能够做到管理有序、进展顺利，其独特之处确实值得各国在开展 e-Science 项目建设时加以学习借鉴。

1. 明确的e-Science建设目标

e-Science 计划的特点是从一开始就提出了比较明确的建设目标。早在 2001 年，英国政府就明确提出发展 e-Science 的目标。即建立一个连接到各区域网格中心的全国 e-Science 中心；建立通用的网格设备及示范项目；网格各学科间的研究协作；e-Science 实验床；参与国际网格项目活动；建立一个网格网络团队。目标是工作的灵魂，有了切实可行的发展目标，保证了英国 e-Science 建设的顺利开展。

2. 合理建设机制——分两个阶段进行建设

最明显的是，英国将整个国家的 e-Science 建设都纳入到国家规划范围内，确定了两个阶段的发展计划。这样的措施有两大好处。一方面，可以统筹规划，避免出现一些重复建设，造成项目重叠，资源浪费；另一方面，可以总结经验，通过前阶段的实施状况，可以将有益的经验在下一阶段中进行推广。

3. 统筹兼顾机制

由于各学科的 e-Science 研究除了有自身的特性外，还有许多共性问题，即对所需要的基础设施有相同的要求。为了避免各学科重复研究，英国制定了 e-Science 核心计划，以建立适于各学科研究的通用技术。英国计划在 3 年时间投资 1200 万英镑建成下一代 IT 基础设施，支持 e-Science 及商业应用，以确保英国在全球网格研究中与美国和欧盟一同处于世界领先地位。

为指导全国的 e-Science 工作，推动实施 e-Science 核心计划。英国在爱丁堡大学和格拉斯哥大学建立了国家 e-Science 中心，目的在于协调和支持十个地区 e-Science 中心，促进 e-Science 技术产业化和商业化。同时负责管理建立于爱丁堡的国家

e-Science 研究院（e-Science Institute）。此外，还负责组织国际 e-Science 交流活动。其目标是力争成为一个全球的 e-Science 研究中心。

英国 e-Science 研究与建设规模可谓盛况空前。截止于 2005 年 8 月 29 日，英国在开展 e-Science 项目研究的过程中，除设立国家 e-Science 中心（NeSC）外，英国在其国内的各个地区、高校设立了地区 e-Science 研究中心，数目达到了十个。为了发挥示范作用，鼓励先进，在此十个地区 e-Science 中心之外，英国还设立了包括贝尔法斯特 e-Science 地区中心在内的七个优秀 e-Science 中心。在强烈的发展意识的驱动下，以强大的资金为后盾，以可靠的科研团队为核心，截止于 2005 年 8 月，英国与 e-Science 相关的项目总数已高达 300 个以上。由于这些地区 e-Science 中心或者优秀的 e-Science 中心均是依靠相关的研究理事会、高校院系、实验室成立，因此，它们完全可以充分发挥这些机构已有的知识积累与学科优势，结合具体的学科与应用领域，个性化地、卓有成效地开展 e-Science 研究与建设。这种统筹兼顾的建设机制，既可以避免重复建设，造成浪费，又有利于发挥个研究理事会、高校院系以及实验室的优势。

4. 双重交流机制

在开展 e-Science 研究与建设的过程上，英国 e-Science 主管部门及科研人员非常注重科研成果的交流与共享，并确定了虚拟交流与实地交流的双重交流机制。

5. 合作机制

e-Science 目标就是实现全球范围内的科研协作。在开展 e-Science 试验研究与初步建设的过程中，各国均非常重视开展各种各样的合作。这种合作不仅包括跨地域的合作，也有跨行业的合作。以英国为例，其 e-Science 核心计划从一开始就注重与工业界的合作，在英国国家 e-Science 中心所公布的“中心工业项目”（中心工业项目一般都是与工业界有一定合作的项目）就达到 49 个。通过与工业界的合作，不仅可以利用他们的资金与技术，更可以将工业界的研发人才吸引到 e-Science 的研究阵容中，最终有利于促进研究上的集思广益。同时也有利于研究成果迅速在工业界得到推广。

在跨地域合作方面，各国也十分注重。例如，英国在其核心计划六大构成要素的第五点中，就突出强调，建设 e-Science，需要建设者“参与到国际网格项目和活动中”。

中国与英国在 e-Science 领域的合作尤其明显。2001 年 10 月，中英第二届科技混委会在京举行。英国科技与创新部部长 Lord Sainsbury 勋爵率领 15 位来自英国政府机构和高校代表出席会议，中国科技部马颂德副部长主持会议，签署了包括在 e-Science 等领域开展合作的会谈备忘录。继中英科技混委会后，英国代表团赴深圳参加高交会。

6. 科学的安全保护机制

针对英国数字科研的项目如洪水般地出现，为方便管理，也为了规范英国国内的各数字科研项目的建设，更为了将责任分摊落实，于 2004 年 4 月英国国家数字科研中心出台了英国 **e-Science** 安全管理政策。该政策从制定背景、制定目的、自我评审、适应范围、人员职责、政策贯彻、制裁等七个方面来确保英国 **e-Science** 项目建设的正常开展。

在该政策的制定目的中，制定者明确指出，制定本政策的目的是，在英国实施数字科研项目的大环境之下，以最完善的实践行动来保证信息安全。

在实施部分，制定者指出：资助项目人员接受相关的培训，以便采用安全方法并在项目伊始就已设计出安全特征；在项目早期阶段就得进行详细的威胁与风险分析；提前制定一份详细的项目安全政策草案，或者还得附上该政策的客观评论；需要进行项目安全审查的特殊地点，并提供资金让审查得到顺利进行；与访问设备、数据集等有关特殊条件；保持特殊安全技术在发展上与时俱进；如果违背这些规定，将会受到制裁。

尽管该政策有支持项目必须考虑安全问题的意图，但是，它也提出了相关制裁措施。假如一个项目肆意或者是因为疏忽而让项目本身或者其他方面的安全面临危险，就有可能受到制裁。即，如果一个项目未能遵照它在获准通知书上所宣称的安全政策，或者未能采用与项目所面临的风险相关的安全方法，该项目就会受得制裁。通过来说，制裁会跟不履行的性质的相匹配。制裁的尺度包括从拒绝访问共享的数字科研设备到撤销已经授予给项目的资源。

7. 公平的评估机制

从 2003 年起，在英国 **e-Science** 指导委员会的监督指导下，英国国家 **e-Science** 中心开始对已经完成的或者正在进行的 **e-Science** 项目实施评估，并出台了若干份评估报告。在评估开展的过程中，除了有网格工程工作组直接参与到与中间件相关的项目评估中，包括各研究理事会在内，不仅对自身承担的项目进行评估并形成评估报告提交英国科技办公室审核，而且还参与了其他没有直接参加的项目评估活动中。在进行项目评估的过程中，研究理事会的各学科委员分别会按领域受理项目评估，参与评估的科学家以“评判员”的身份给每份项目报告提出评估意见。纵观英国 **e-Science** 项目的评估过程，不难发现，及时、公正、公开、公平是它的出色之处。

8. 与特定学科和应用领域相结合

可以说，**e-Science** 能够在短时间内“火爆”起来，并非一种偶然。目前，从各国的 **e-Science** 项目的进展现状，尤其是英国各 **e-Science** 研究中心的现状分析，**e-Science** 的研究与建设基本上都与特定学科相结合。如英国粒子物理学与天文学研

究委员会在 e-Science 的研究与开展过程中,就密切结合了天文学;牛津地区 e-Science 中心则将 e-Science 与工程科学结合到一块进行研究,等等。以英国为例,各个中心均是结合一定的学科研究来开展 e-Science 的。

除了与学科结合进行研究,英国在开展 e-Science 的过程,非常注重将 e-Science 的研究与实际应用相结合。目前,e-Science 的应用领域包括但不限于:粒子物理、乳房造影、重要系统的模拟和整合、超级计算中心、重要科学仪器的共享、天体物理学、生物和生物信息学网格、虚拟实验室等。

9. 网格技术是核心支撑

网格技术集成了计算机科学领域中多种技术的最新成果,代表了信息技术和信息化发展的方向,特别是在现阶段,突出体现和满足了科学研究的需要。因此,在很多场合,网格几乎就是 e-Science 的代名词。正是通过网格才有可能将 e-Science 所需的众多技术与资源,如分布式计算技术、网络安全技术、协同工作技术、资源管理技术等等集成一体,构成并实现了 e-Science。

在英国的 e-Science 建设与研究项目中,有超过一半的项目与网格直接相关,另外的一些项目也与网格在不同程度上存在关系。这一方面说明 e-Science 的顺利实现与推广离不开网格,另一方面也是说明,英国在发展 e-Science 的过程中,始终是以网格技术作为 e-Science 的核心支撑。

10. 各承担主体均有清晰的组织管理体系

目前,在英国 e-Science 的发展格局中,参与 e-Science 项目建设的主体,既有研究理事会,也有新兴的 e-Science 中心。在项目开展的过程中,这些承担主体无一例外地建起清晰的组织管理体系,从而确保了 e-Science 的顺利开展。

例如,国家 e-社会科学中心在成立伊始,它就设立执行主任一名,副执行主任两名,研究主任一名,中心秘书一名,e-社会科学、数据归档研究带头人一名;行政管理人员一名;学习技术员一名,项目经理一名;网格工程师两名。而国家文本挖掘中心则设立中心设主任一名,副主任三名,主任助理两名,商业经理一名,负责项目一名。

开放中间件基础设施研究所(Open Middleware Infrastructure Institute, OMII)在这方面也比较有特色,为保证各项工作的顺利开展,该研究所设置了如表 2-8 所示的组织管理体系。

当然,英国 e-Science 之所以能够取得如此巨大的资金投入并在短时间内取得迅速发展,离不开英国政府对 e-Science 的高度重视,2001 年,英国政府将 e-Science 列为 2001—2004 年度三个重点研发领域之一。同时,英国 e-Science 研究从起步阶段就开始考虑工业界的利益,注意 e-Science 研究成果的商业应用价值。

表 2-8 开放中间件基础设施研究所的组织管理体系

所长、项目负责人、执行主任	Prof. Peter Henderson
项目第二负责人	Prof. David De Roure
商业发言人	Dr. Alistair Dunlop
副主任	Dr. Steven Newhouse
首席架构师	Prof. Syd Chapman (IBM UK)
商业顾问	Dr. Colin Upstill (IT Innovation)
技术顾问	Dr. Mike Surridge (IT Innovation)
	Dr. Tony Storey (IBM UK)
质量工程师	Mrs. Claire Walker
高级研究员\工程师	Dr. Juri Papay
	Dr. Gary Li
	Dr. Brian Carpenter
	Mr. Justin Bradley
	Dr. Stephen Crouch
研究助理/研究工程师	Dr. Karen Ng
网站负责人	Mr. Victor Chang
	Ms. Jodi Crisp
系统管理员	Mr. Aaron Wookey
临时秘书/行政管理人员	Ms. Lucy Bartram

第 3 章 美国e-Science发展现状分析

首先需要指出的是，e-Science 这一名词已经不仅仅局限于英国始于 2001 年的一项耗资巨大的计划名称，而更为广泛接收的是其基于先进的信息技术、计算机技术所实现的科研合作的本质，和其为各种科学领域的研究活动提供的一种环境、平台、基础设施及技术保障，换句话说，就是 e-Science 在应用和技术两个层面上的内涵。也正是沿着这两条主线。

3.1 美国e-Science历史背景

3.1.1 技术背景分析

在英国的 e-Science 计划中，网格被选定为实现目标的依托技术。对于这一事实，当剥去网格技术符合科学研究统一协调管理的表层解释外，看到的是 e-Science 对于基于计算机网络的分布资源调度和资源与服务共享的深层次需求，而这也恰是其技术支撑平台的本质和面临的问题之所在。但无论是分布资源调度，还是资源服务共享，或者是计算机网络技术，美国无疑都有着丰厚的历史，并占据着不容忽视的地位。从各种概念的提出、技术的实现，到协议标准的制定以及应用产品的开发与使用，美国似乎总有一支系统化的科研队伍在不遗余力的做出贡献，有杰出的学科领袖，有强大的研究机构、组织和团队，有分布缜密的项目计划，更有使用充沛的资金注入和性能良好的硬件环境。

具体到网格技术，美国毫无例外的具有着绝对的发言权，从五层沙漏结构到 Open Grid Services Architecture（以下简称 OGSA）的提出，从 Open Grid Services Infrastructure（以下简称 OGSi）到 Web Services Resource Framework（以下简称 WSRF）的应用，从 Condor 到 Globus，从 20 世纪 90 年代中期的 FAFNER（Factoring via Network-Enabled Recursion）和 I-WAY（Information Wide Area Year）再到现在的 Access Grid 和 TeraGrid，都在用事实更新着我们对于这一技术的认识。另外，美国在网格方面的研究优势还包括拥有像“网格之父”Ian Foster 这样的领军人物，和来自高校、国家实验室及其他科研机构的实力强大的研究团队，以及来自于如 National Science Foundation（以下简称 NSF）和 Department of Energy（以下简称 DOE）等政府机构或民间组织大力的资金支持等。可以说，目前美国已经为其进行科研合作构筑了一个坚实并不断发展的基础设施平台。

3.1.2 应用历史分析

早在希腊时期，科学家们就已经开始了某种程度的交流活动，只是那时的科学研究还只局限于爱智者对知识的探索。文艺复兴、工业革命，尤其是进入二十世纪以来，近现代科学进入了“大科学”时代，更多需要的是一个领域，或者社会，甚至全球的共同协作的大系统。在这种大的环境背景下，美国作为科学研究活动最为繁荣的国家之一，也面临着同样的机遇与挑战。其实，在阿波罗登月计划以前美国就已经开始了大规模的科研合作，只是随着计算机网络的发展，信息技术、计算机技术的进步，这种合作有了更加有效和功能强大的保障，而且，应用领域和范围也在发生着多方位的扩展。比如，物理学方面有 Grid Physics Network（以下简称 GriPhyN）、Particle Physics Data Grid（以下简称 PPDG）和 U.S. ATLAS，化学方面有 Collaboratory for Multi-Scale Chemical Science（以下简称 CMCS）和 Research Surge Enabled by Cyberinfrastructure（以下简称 RESURGENCE），天文学有 U.S. National Virtual Observatory（以下简称 NVO）和 TeraScale Supernova Initiative（以下简称 TSI），地理学有 Network for Earthquake Engineering Simulation（以下简称 NEESit）和 Community Climate System Model（以下简称 CCSM），生物学有 Biomedical Informatics Research Network（以下简称 BIRN）、Science Environment for Ecological Knowledge（以下简称 SEEK），另外，还有航空航天领域的 Information Power Grid（以下简称 IPG）和科学模拟领域的 Terascale Optimal PDE Simulations（以下简称 TOPS）等，可以说美国的 e-Science 目前已经涉及到了科学研究的各个角落。而在这些系统全面地通过计算机网络，利用先进的信息技术、计算机技术进行的各学科的科研合作的背后，却没有一个类似于英国 e-Science 计划这样的统一管理机制，取而代之的是分散在 NSF、DOE 或 NASA 等机构中很多资助计划的某些部分的有机组合，也就是说美国的 e-Science 项目是在需求和针对需求的具体实现的基础上，慢慢聚合到了一起，从而形成了今天的规模，这不得不再次提醒我们 e-Science 科研的本质，并向我们展示 e-Science 的产生与发展是科研活动的需要，是大科学的需要，是生产力发展的需要，更是我们这个时代的需要。

3.2 美国e-Science项目分析

下面将从具体的项目出发，汇总美国 e-Science 项目的资助结构、承担结构和整体的体系结构，并对项目的类型和研究领域展开进一步分析。

3.2.1 项目选取方法说明

因为美国没有明确提出 e-Science 的计划，而相关的项目是分散在 NSF、DOE 等机构的资助计划中，所以给选取调研项目带来了一定的困难。具体的解决方案如下：

1. 确定选取标准

- ① 依托计算机网络，使用信息技术或计算机技术进行的科研合作或资源服务共享，或者是有关的技术研发；
- ② 具有一定的典型性和代表性；
- ③ 有独立的网站或至少有介绍性的网页；
- ④ 需要有较完备的项目信息（至少要包括项目名称和适当介绍）。

2. 选取方法

- ① 利用多种渠道，如搜索引擎等，初步查找到一批项目网站或相关文献；
- ② 使用追溯法检索，通过文献的引文、网站链接等进一步查找项目；
- ③ 通过已有项目归纳母项目、资助机构、承担机构或个人，再以此为出发点进一步进行查找。

3. 问题

- ① 遗漏：可能查到的项目过多的汇聚于某些领域，而遗漏其他领域的项目；
- ② 不相关：从已有项目归纳出的结果做进一步查找时会出现很多不相关的项目使用引文或网页链接查找时，项目的相关度会随着原始资料的引用距离的增加而下降。

通过以上方法，最终选取美国调研项目 94 个，国际合作项目 17 个，列表参加附录 1。因为本次调研所选取的 e-Science 项目是建立在第一轮通过各种途径大范围搜集的基础上的，因此可以说取得的样本是具有一定的代表性，可以通过对其进行的分析一窥美国在 e-Science 领域的发展情况。

3.2.2 资助结构分析

在调研的 111 个项目中，共涉及到 63 个资助机构和计划，其中除了来自美国地区的资助外，还包括 14 个来自其他国家的资助机构。美国地区资助机构的类型主要包括政府部门、高校研究中心、公司企业、图书馆、研究机构和州立机构。资助项目数量见表 3-1。

表 3-1 调研项目中涉及的美国地区 e-Science 项目资助机构

美国地区资助机构	接收资助项目数量
U.S. Department of Energy	57
National Science Foundation	40
Department of Defense	6
National Aeronautics and Space Administration	5
Department of Health and Human Services	4
San Diego State University	4
University of California,San Diego	3

续表

美国地区资助机构	接收资助项目数量
University of Wisconsin-Madison	1
United States Patent and Trademark Office	1
Ohio Board of Regents	1
National Oceanic & Atmospheric Administration	1
National Library of Medicine	1
National Archives and Records Administration	1
Microsoft	1
IBM	1
Cisco	1
Library of Congress	1
University of California, Los Angeles	1
California Institute for Telecommunications and Information Technology	1
California Digital Library	1

从表 3-1 中我们可以发现，美国地区的 e-Science 项目主要还是由政府部门，尤其是 NSF 和 DOE 两大科学研究资助机构进行支持的，这主要是因为 e-Science 虽然有着广阔的发展前景，对科学研究、对社会甚至整个人类文明的进步也将起着极大的推动作用，但毕竟这是一项耗资巨大、牵扯多方因素的工程，而且不会在短期内收到更多利益上的回报，对于这样的投资大、风险高、见效慢，但却关系国家利益的项目，政府机构是责无旁贷的投资者，而那些感兴趣的民间组织，或想要回报社会的公司企业或者个人，也许也会积极参与，但在整个的 e-Science 发展中，这种资助绝对不会占据过高的份额。另外一点我们可以从表 3-1 的数据中发现的问题是，在 e-Science 项目的资助机构中，除了科研部门外，还涉及一些其他类型的部门，如美国国防部、美国卫生与人类服务部、美国专利和商标事务所、国家海洋和大气管理局、国会图书馆等，对于这一现象的一个解释是 e-Science 已经进入了人类社会的方方面面，但更为具体的原因是因为 e-Science 的底层技术研究不仅仅可以用来进行科研合作，而是可以在更为广泛的领域应用，比如电子商务、信息服务、国家安全、教育培训、卫生医疗等，因此会吸引相关的机构对此进行投资。图 3-1 显示了美国对于 e-Science 项目投资的机构和项目之间的隶属关系。

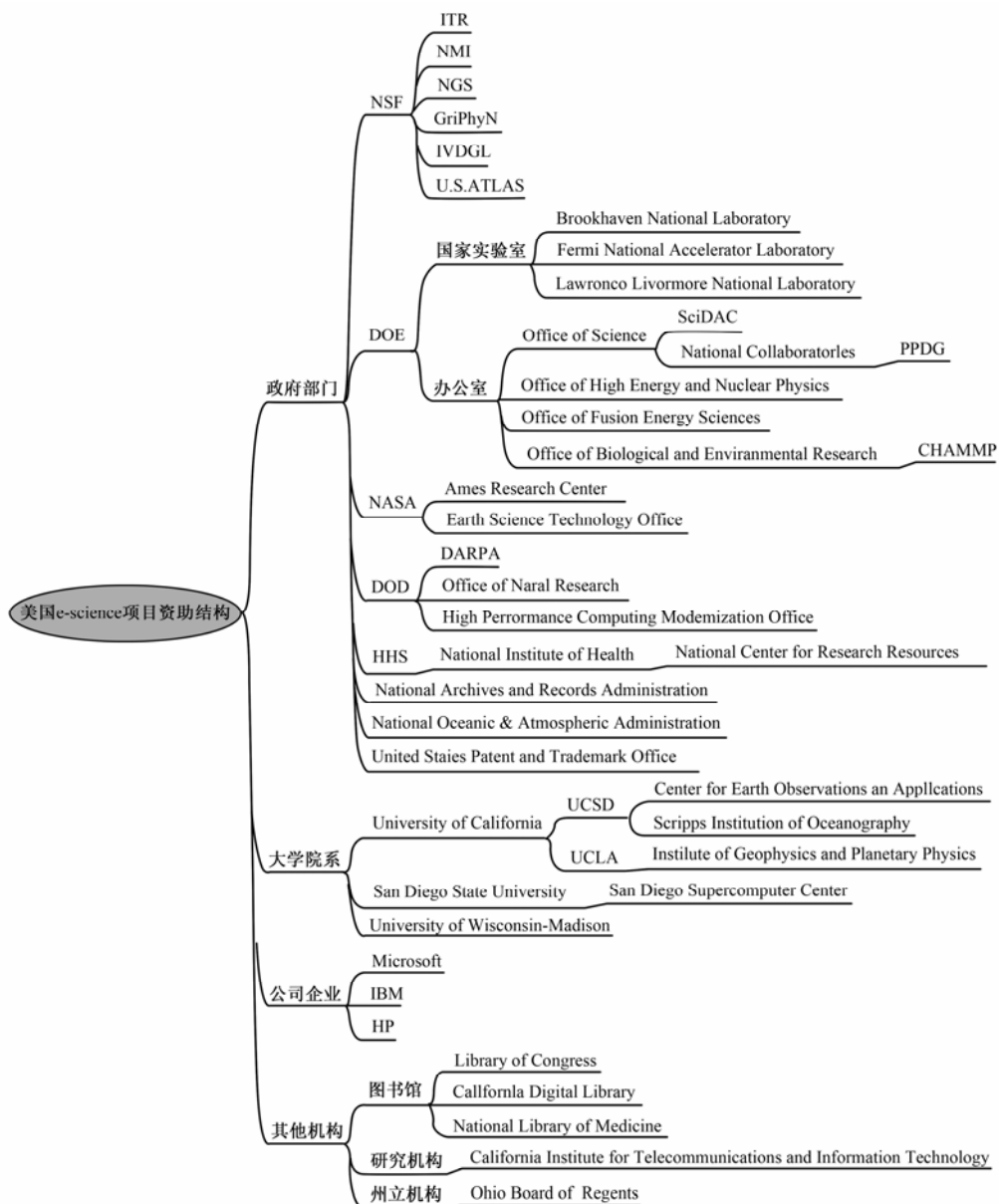


图 3-1 美国 e-Science 项目资助结构

3.2.3 主要资助机构和资助计划介绍

通过表 3-1，以及对已调研的项目的资助信息进行的分析，可以总结出主要的 e-Science 资助机构及其资助计划，下面就选取其中较为典型的 NSF 和 DOE 的 4 个资助计划做进一步的介绍和分析。

1. NSF

作为美国科学研究的主要资助机构，NSF 无疑是大量 e-Science 项目主要的资金来源之一，在调研的 111 个项目中，有 37% 的项目都接收过 NSF 的资助。NSF 涉及 e-Science 项目的主要资助计划包括：

(1) NSF Middleware Initiative（以下简称 NMI）

NMI 计划始于 2001 年，其主要目标是开发、部署并维护一套可重用、可扩展的中间件服务，以为科学、工程应用提供一个网络化的环境，从而通过数据、设备、计算资源共享提高科学生产力，并促进研究与教育的合作。截止于 2005 年，该计划共包括子项目 54 个，资助金额 4145.4531 万美金，本次调研涉及其中 6 个项目。NMI 的资助范围包括：中间件部署（Deployment）、中间件开发（Development）、探索性研究（Small Grants for Exploratory Research, SGER）、专题讨论会（Workshop），具体的资助结构如图 3-2 所示。由图 3-2 可分析出，NMI 计划的资助重点在于中间件的开发和部署方面，但不论是中间件的开发还是部署，都是属于对 e-Science 支撑技术方面研究的资助。

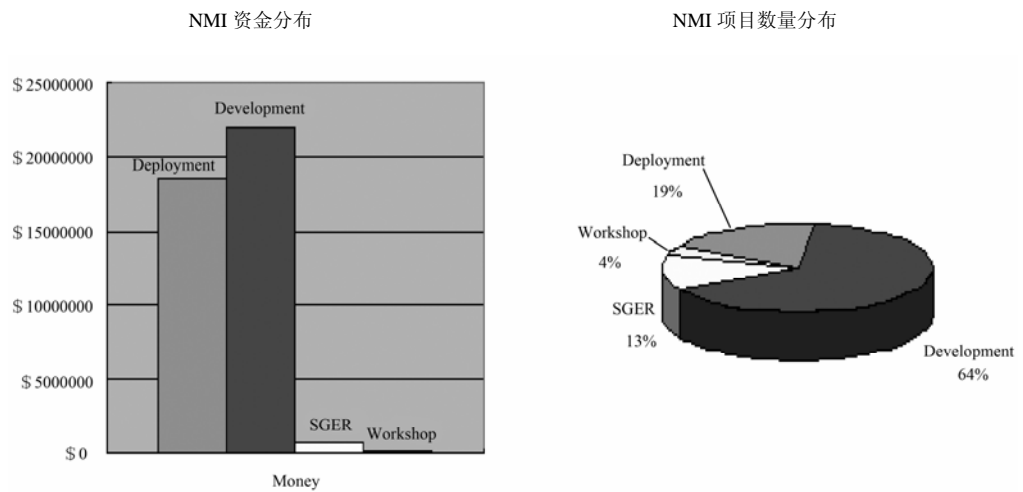


图 3-2 NMI 资助结构

(2) Information Technology Research for National Priorities (下文简称 ITR)

NSF 的 ITR 计划始于 2001 年, 重点在于支持那些对使国家处于世界领先地位的研究起支撑作用的信息技术研究。其中使国家处于世界领先地位的研究包括: 科学与工程领域的先进技术研究; 发展生产, 繁荣经济和活跃文明社会的研究; 国土安全方面的研究。

ITR 的资助金额共 13244.5737 万美金, 包括 142 个项目, 内容包括: 计算、网络、人机界面和支持可靠、复合的分布系统的信息管理的整合; 整合数据、模型、通信、分析和/或控制系统, 包括动态的、数据驱动的在预测、风险评估和决策制定方面的应用程序的各种革新方法; 信息系统与社会系统的交互及复杂的相互依赖关系; 计算模型的创新或研究教育模拟。

这些项目包括对材料研究、电信、工程与计算机科学、工业技术、海洋学、计算机科学、课程编制、人力资源与培训、人文学科、软件开发、数据银行与软件设计、信息系统和其他应用等领域的研究, 具体的项目资金分配和项目数量的领域分布情况如图 3-3 所示。

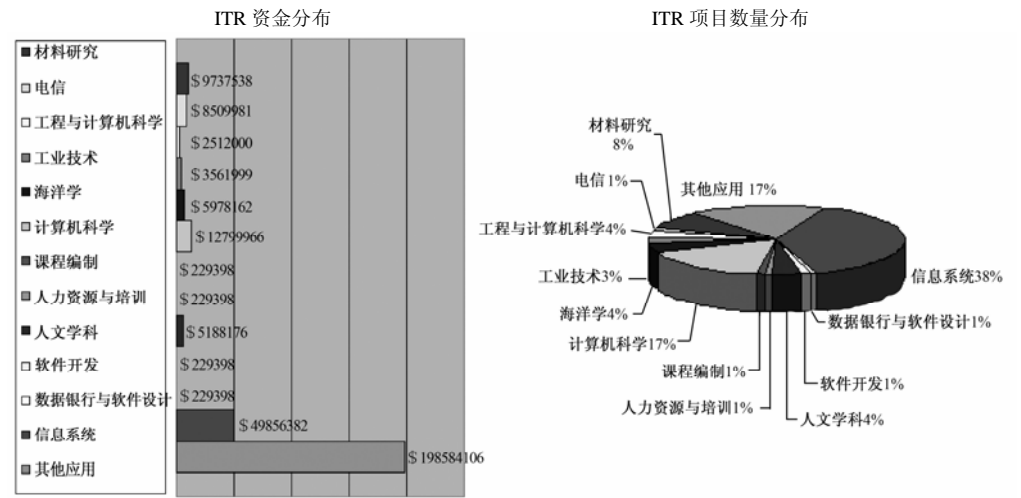


图 3-3 ITR 资助结构

由图 3-3 可以看出, 在 ITR 计划中, 虽然对于技术的研发仍然占据着很大的比重, 但更多的资金与项目是针对具体研究领域的应用的。

(3) NSF 的其他计划

其实在 NSF 的众多研究资助计划中, 除了上述提到的两个典型代表外, 还有很多也包括了 e-Science 项目的资助, 如“下一代软件”(Next Generation Software, 以下简称 NGS)计划, 其中就包括了很多对 e-Science 基础技术的研究。另外还有针对具体学科的资助计划汇总, 也包含了很多通过计算机网络、利用先进的信息技术和计算机技术进行科研合作的项目, 但由于这些资助计划中的 e-Science 项目过于零散, 所以这里就不再一一赘述。

2. DOE

美国能源部是美国 e-Science 项目的又一主要的资助机构，尽管其主要资助物理学相关领域的研究，但对于其他学科，如生物学、环境科学、材料科学、化学和计算机科学等也给予了大量的研究经费的支持，并设立了很多相应的研究资助计划。其中，科学办公室（Office of Science）可以说是 e-Science 项目资金的主要来源和管理者，在本次调研选取的项目中，有近半数的项目接收过其资助。另外，由科学办公室管理的 10 个美国国家实验室，也是 e-Science 项目的主要承担机构。

DOE 对于 e-Science 的主要资助计划包括 Scientific Discovery through Advanced Computing（以下简称 SciDAC）计划和 National Collaboratories 计划。

（1）DOE SciDAC

2000 年，美国国会拨款近 6 亿美元给 DOE，批准了这项耗资巨大的研究计划。2001 年，SciDAC 计划正式启动，其主要目标正如该计划的名称所说的那样，“通过先进的计算实现科学发现”。该计划旨在建立科学计算的软件、硬件框架，以开发大规模计算在先进科学发现和工程应用方面的潜能，在数学、计算机科学与物理、化学、生物学、环境学中的计算问题间飞架一座桥梁，使天堑变成通途。SciDAC 主要关注 3 个方面的活动：数学模型、计算方法和利用大规模计算机进行的科学应用；开发计算系统软件以加速科学代码的开发，从而实现大规模计算机最大的有效性，并使大范围的科学家能够在其试验中使用计算模拟；开发网络技术、合作工具以链接 DOE 科学办公室的计算资源，来支持那些在地理上分散的科学家，方便大规模数据集的移动。

到 2006 年 1 月初，SciDAC 共资助了 69 个项目，分布于 5 个研究领域，具体包括：基础能源科学、生物与环境研究、聚变能科学、高能物理与核物理、先进科学计算研究。其中在先进科学计算领域的研究又细分为：计算机科学组合软件框架、数学组合软件框架、合作与网络 3 个方面。SciDAC 的具体的资助结构如图 3-4 所示。

从图 3-4 中我们可以看到，不论是从资金分配角度，还是从项目数量发布角度来说，SciDAC 都更为关注底层计算环境的研发与部署。而在具体应用研究方面，SciDAC 仍然表现出 DOE 关注能源与物理研究的特色，对与此有关的科研项目给予了更多支持。

（2）DOE National Collaboratories

始于 2000 年的 DOE National Collaboratories 是 DOE 科学办公室在对大规模科研合作的支持方面进行的又一项投资。该计划旨在研究、开发和部署分布式的系统软件和合作框架，以连接在地理上分布的科学研究人员，从而存进大规模数据集流动，并保证科学家和 DOE 的实验室能够加入到一个大规模的分布科学研究之中。目前，该计划的研究子项目共 26 个，涉及 4 个研究范围，具体包括：合作示范项目、合作框架研究、合作中间件研究、高性能网络研究。其项目数量的分布情况如图 3-5 所示。

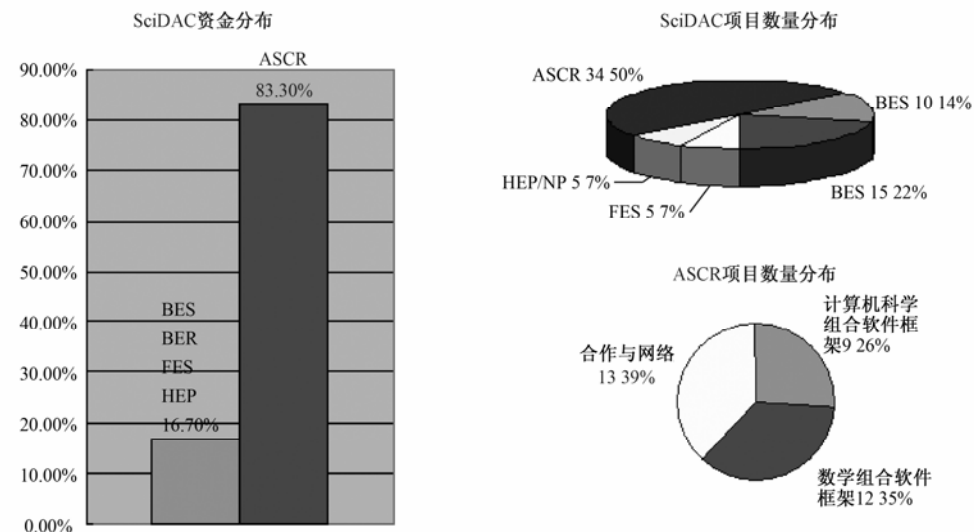


图 3-4 SciDAC 资助结构

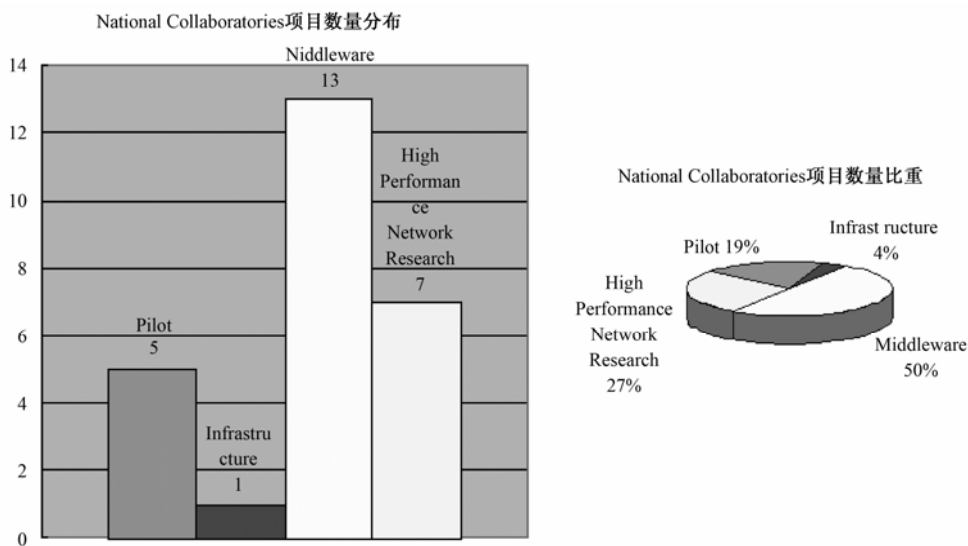


图 3-5 National Collaboratories 项目结构

由 National Collaboratories 计划的项目分布，可以看出，对于科研合作所涉及的基础、中间件相关技术，以及底层的网络的研究构成了该项计划研究的主体，而应用示范项目只占到了整体计划的不足 1/5 的数量。这些应用示范项目所涉及的具体应用学科领域包括：地球科学（Earth System Grid II）、高能物理（Particle Physics Data Grid Collaborative Pilot）、聚变能研究（National Fusion Collaboratory）、化学

(Collaboratory for Multi-Scale Chemical Science) 等。

其实, 美国对 e-Science 项目的投资实际涉及到了很多机构和很多资助计划, 本次调研实际只能做到对典型机构和典型计划所资助的典型项目进行搜集和整理, 并加以分析介绍。经过分析和总结, 之所以造成 e-Science 项目的资助机构和计划如此分散和广泛, 其根本原因是由 e-Science 的本质决定的。通过前文的分析, 我们知道 e-Science 的研究主要涉及两个方面: 具体的应用和底层技术的开发, 而其中的应用部分, 实际又是在底层技术平台的基础上, 具体学科领域藉此实现各自特定的目标。那么在这一过程中便产生了两个 e-Science 项目的扩展点:

底层技术主要是支持合作和共享的, 而在这两个领域所涉及的网络技术、信息技术、计算机技术可以说是多如繁星, 网格技术可以说只是其中一点。另外, e-Science 的技术研究又不局限于此。因为 e-Science 根本上是要为科学研究提供所有的技术支持, 这就不仅仅是分布计算、资源服务共建共享那么简单, 比如还要涉及建立在分布环境上的应用软件框架的设计或具体应用的开发, 涉及基于大规模计算机实现的科学模拟的实现和应用等, 这就无疑扩充了 e-Science 技术研究的范畴。

具体学科领域的应用是造成 e-Science 研究项目数量繁多、范围广泛的另一个根本原因。e-Science 的本质仍然是科研, 与传统科研的不同点笔者认为主要是在于利用了技术手段实现了大规模、大范围的科研合作, 从而解决了以前不能或很难解决的问题。而在网络、信息技术、计算机技术蓬勃发展的今天, 利用这些技术完成各种工作已经是一种不可逆转的趋势, 也就是说随着时间的推移, 会有越来越多的科研活动会借助这些先进的技术所实现的强大功能, 那么根据 e-Science 的定义, 这些也就都被囊括进了 e-Science 的体系中去, 使得这套系统愈发庞大与复杂。

但无论如何, 在美国, 科研工作者与管理者们建设 e-Science 的步伐并没有因为没有一个统一的称谓而放慢, 相反地, 一套自底向上, 从需求出发的利用先进技术实现的科学研究体系正慢慢地在各种资助机构和资助计划的引导下, 显示出其不俗的实力, 为人类社会创造着不朽的财富。

3.2.4 项目承担机构的构成分析

本节将根据调研得到的数据, 对美国 e-Science 项目的承担机构和相应的研究人员进行分析。

1. 承担机构分析

在调研的 94 个美国地区 e-Science 项目中, 共包括 160 个项目承担机构, 如果对美国 e-Science 项目承担机构按如下类型进行划分: 政府部门、科研机构、高校院系、研究中心或实验室、公司企业、协会组织、其他机构 (如图书馆), 则承担机构的数量和所承担的项目数量按类型分布情况如图 3-6 所示。

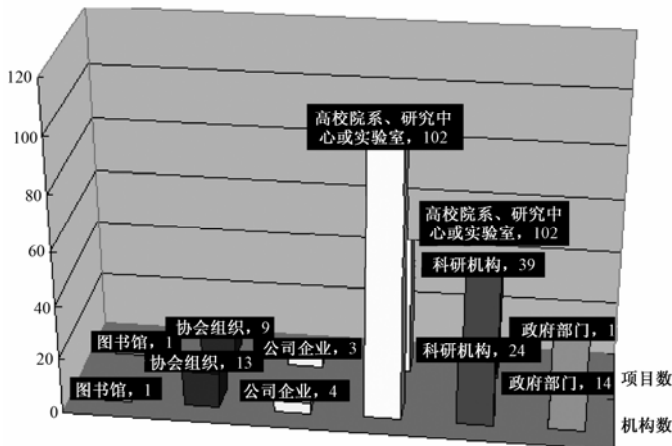


图 3-6 承担机构数量及所承担项目数量按类型分布

从图 3-6 和图 3-7 我们可以看出，尽管高校在数量上占有优势，但是其平均机构承担项目数却是最低的。造成这一现象的主要原因是，在高校这一类型中包括高校院系、高校研究中心和高校实验室 3 个部分，其中只有高校研究中心和高校实验室，以及少量的高校院系具有较强的研发能力，能够在多个 e-Science 的研究项目中承担主要的研究任务。而剩余的，也是在高校类型的承担机构中占较大比重的是研发能力较弱的高校院系，这些高校院系可能只是偶尔在某些特定领域的项目研究中作为项目应用的参与者出现。另外，e-Science 本质决定着其研究项目，尤其是应用领域的研究项目需要很多机构的共同合作，再加上 e-Science 研究本身的先进性和项目资金的充裕，于是吸引了大量的高校院系参与其中，而最终导致高校平均承担项目数量不多的结果。例如，仅 U.S. ATLAS 一个项目就包括了 30 个高校，而其中有半数以上仅参加了该项目的研究。

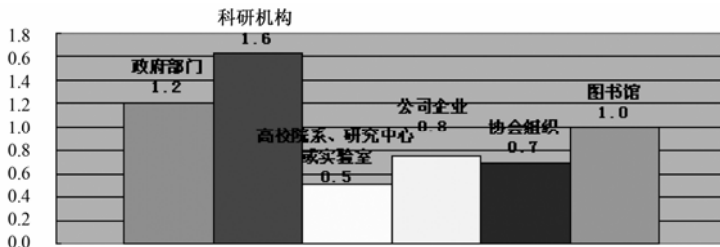


图 3-7 不同类型承担机构平均承担项目数量

表 3-2 是承担项目数超过 3 个（不包括 3 个）的承担机构列表。承担项目数量越多，在某种程度上至少可以说明该承担机构在 e-Science 相关领域的研究方面具有一定的科研实力。通过表 3-2 我们可以看到，由 DOE 科学办公室管理和资助的 17 个国家实验室中有 12 个出现在此次调研中，在资助项目数量排名前 10 位的机构中就包

括了 6 个。这一方可能是由于本调研选取了大量 DOE 科学办公室资助的项目，但从另一方面也在某种程度上说明了这些国家实验室在 e-Science 研究部署与应用过程中扮演了较为重要的角色。当然除了这 12 个国家实验室，还有很多科研机构 and 高校也是 e-Science 项目的重要参与者，如印第安纳大学、圣地亚哥超级计算中心、南加州大学信息科技学院等。

表 3-2 承担 3 个以上项目的机构列表

美国 e-Science 项目承担机构	承担项目数量
Argonne National Laboratory	25
Lawrence Berkeley National Laboratory	23
美国 e-Science 项目承担机构	承担项目数量
Indiana University	12
Lawrence Livermore National Laboratory	12
Oak Ridge National Laboratory	12
San Diego Supercomputer Center	12
Brookhaven National Laboratory	10
University of Southern California Information Sciences Institute	10
Los Alamos National Laboratory	9
National Center for Supercomputing Applications	9
University of Chicago	9
University of Wisconsin at Madison	9
Pacific Northwest National Laboratory	8
University of California at San Diego	8
Massachusetts Institute of Technology	7
Fermi National Accelerator Laboratory	6
Sandia National Laboratory	6
National Center for Atmospheric Research	5
Stanford Linear Accelerator Center	5
University of Utah	5
California Institute of Technology	4
General Atomics	4
Harvard University	4
State University of New York at Stony Brook	4
University of Arizona	4
University of California, Los Angeles	4
University of Illinois at Urbana-Champaign	4
University of Maryland	4
University of Southern California	4
University of Tennessee	4

2. 研究人员分析

科研人员是 e-Science 项目的重要组成部分，对于研究人员的分析可以从多个角度进行，但由于调研数据所限，目前只能从参与研究人员的数量角度对这些项目进行大概分析。

对于一个项目，研究人员的数量还是能够从一个侧面反映项目的规模水平。按照 30 人以下、30~100 人、100 人以上 3 个等级，此次调研的美国地区 e-Science 项目按研究人员数量的分布情况如图 3-8 所示。

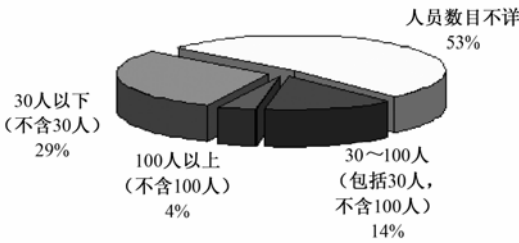


图 3-8 美国 e-Science 项目按人员数量分布

通过数据还可以发现，研究人员超过 30 人，尤其是 100 人的项目中，大多数都属于对 e-Science 应用领域研究。究其原因，主要是由 e-Science 面向大规模、大范围科研的特点决定的。既然是要解决科学合作的问题，当然就要在实际的合作过程中对问题进行发现和求解，因此必然会涉及多个承担机构，从而也就会涉及相对更多的研究人员。也正因此，我们需要更加客观地评价研究人员数量在项目评价方面的意义，尤其是对 e-Science 这类项目的评价。举个简单的例子，Virtual Data Toolkit 项目的研究人员只有 23 个，圣地亚哥超级计算中心的 Storage Resource Broker (SRB) 项目的研究人员只有 17 个，即使是 Condor 项目，其研究人员也只不过 38 个，但我们知道这些项目的研究成果都是 e-Science 支撑技术平台中重要的组成部分。

另外，通过对调研项目负责人数据的分析，我们还可以根据其所负责项目的数量列出几个美国 e-Science 领域的专家，如 Ian Foster、Miron Livny、Al Geist 等。

3.2.5 项目组成分析

1. 应用领域

从整体上，e-Science 项目可以分为应用型研究及建设和底层关键技术研究，其中后者将在“技术框架与技术模块”中给予详细介绍和说明。

首先调研的美国地区项目共包括应用型项目 29 个，分布于物理、化学、天文学、地球科学、生物学和工程应用等领域，再加上底层技术研究，则具体分布情况如图 3-9 所示。

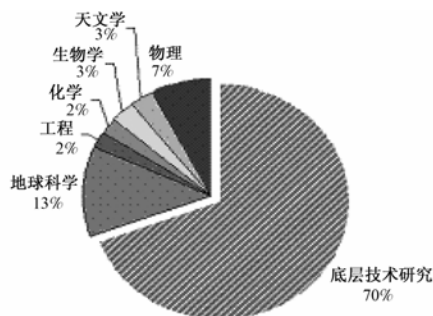


图 3-9 美国 e-Science 项目应用领域分布

从图 3-9 我们不难发现大量项目正在从事着 e-Science 技术方面的研发，这从某个侧面也反映了美国对 e-Science 底层技术的重视程度。而在应用层面上，地球科学和物理学占据了较大的比重，其中在地球科学方面的研究包括大气科学（主要涉及气候、气象研究）、地质学（主要涉及地震研究）以及地球观测等领域；而在物理学方面，则以高能物理的研究于试验等领域为主。分析结果表明，项目较为集中的应用领域，正是那些需要大量科研人员参与、进行复杂或大范围的科学试验、产生大量数据的科学研究，是那些单凭少数科学家或单个研究机构很难实现的科学研究，换句话说，就是那些需要大规模、大范围合作的科学研究，是“大科学”，而这也正是 e-Science 的目标所在。

2. 国际合作项目

通过对调研数据统计分析，美国共在 22 个 e-Science 项目中与 37 个国家和地区进行合作，其中有 5 个项目中美国是作为项目的发起者和领导者。如果以项目数作为合作次数进行计算的话，则各国与美国在 e-Science 的研究建设过程中的合作情况如表 3-3 所示。

表 3-3 美国 e-Science 国际合作情况

合 作 国 家	合 作 次 数	合作国家和地区	合 作 次 数	合作国家和地区	合 作 次 数
英国	15	俄罗斯	2	芬兰	1
荷兰	5	捷克	2	克罗地亚	1
加拿大	5	挪威	2	马来西亚	1
澳大利亚	4	欧洲	2	南斯拉夫	1
韩国	4	泰国	2	葡萄牙	1
日本	4	新加坡	2	瑞典	1
中国	4	意大利	2	斯洛伐克	1
德国	3	中国台湾	2	委内瑞拉	1
法国	3	阿尔及利亚	1	西班牙	1
瑞士	3	爱尔兰	1	新西兰	1
印度	3	奥地利	1	智利	1
巴西	2	波多黎各	1		
波兰	2	菲律宾	1		

其中作为 e-Science 项目的发起国,英国与美国在相关领域的合作最为频繁,在美国所有的国际合作项目中,有 68.2% 的项目都有英国的参与。中国(包括台湾地区)与美国共在 4 个项目上进行了合作,分别是 Solenoidal Tracker At RHIC experiment(以下简称 STAR)、Pacific Rim Applications and Grid Middleware Assembly(以下简称 PRAGMA)、ACCESS Grid 和 BaBar,其中的 STAR 和 BaBar 均是高能物理方面的 e-Science 应用项目。

3.3 美国 e-Science 技术框架与组件

尽管网格技术不是实现 e-Science 的唯一可选项,比如在分布资源调度方面,P2P 就为人们所津津乐道。网格技术虽然强大,但却不可能解决所有大规模科学研究中所遇到的技术问题,比如网络连接问题和基于分布式环境的应用软件的设计与实现问题等。另外,网格技术本身也有着各种各样的不足和限制,其与 Web Services 的融合,对语义网技术的使用等事实都充分说明了这一点。但不容否认的是,网格技术确实能够为 e-Science 提供了一种可以实现的框架,通过为发现和导航对远程资源的访问,提供一种可伸缩的、安全的、高性能的机制,网格使空前规模的科研合作成为了可能。因此,本节将对 e-Science 的技术框架和功能模块的研究重点放到对美国网络的分析上。

3.3.1 网络框架

按照英国国家 e-Science 中心对于 e-Science 框架所定义的 31 个组件对美国的 e-Science 项目进行描述。在这一框架中,共分为 3 个层次,31 个组件,具体的项目分布情况见表 3-4,其中项目数未去重。

其中应该指出的是,尽管英国的这套网络框架没有列出构造层和连接层的内容,而我们的调研项目也很少涉及这部分研究,但它们仍然是整个 e-Science 更为底层的基础,是在建设 e-Science 的过程中不容忽视的部分。尤其是其中的网络连接问题,在网格五层沙漏结构中被描述为瓶颈之一,而实际情况也恰恰如此,尽管目前的网络连接已有了飞速的进步,但面对大规模科研对大量数据传输、远程控制、实时数据分析等高水平的需求,仍然需要给予很大的重视,比如美国 NSF-ITR 中的 UltraLight 项目,就是在提高网络性能方面做出的努力。有学者认为,网络框架主要包括资源层、服务层、应用层和问题求解环境。

(1) 资源层

包括各种分布异构的可以为 e-Science 使用的资源,如计算机、数据存储系统、仪器设备等。其中 Condor 提供了在工作站池中管理资源的机制。另外,还需要对该层资源进行驻留和监测的整合,以向用户提供稳定的服务。

（2）服务层

在服务层，网格提供通用的服务和统一的资源接口。主要的服务包括授权认证服务、资源访问服务、网格信息服务等，另外，还提供网络监控、记账、错误管理等服务。具体可以参见表 3-4 中的服务。在这一层中，集中了大量的研究项目，并已经取得了很大的进步。

（3）应用层

在这一层中，我们可以看到大量的应用于不同领域的工具。例如可视化工具、合作交流工具、数据管理工具等。

（4）问题求解环境

在网格框架的顶层，为人们提供了一个组织问题解决方案的框架和机制。

总之，这些不同的服务和工具呈现出了一个稳定的技术基础，尽管还很不完善，但却已经可以支撑 e-Science 的开发、部署和应用。

表 3-4 美国 e-Science 项目技术研究分布

所属层次	组件名称	项目总数
第三层（63）	应用服务	31
	网格计算服务	31
	领域本体、任务本体与元数据	13
	发布与归档服务	0
	问题求解环境	5
第二层（69）	合作工具包	13
	数据管理服务	32
	数据挖掘与分析服务	20
	网格管理服务	6
	仪器管理服务	0
	市场经济服务	1
	模式与程序原型	11
	时序安排、监测与诊断服务	1
	虚拟与沉浸环境服务	4
	可视化服务	14
	工作流服务	14
第一层（62）	审核	2
	认证与授权	14
	代理	2
	合作与远程仪器服务	17
	协同调度	6

所属层次	组件名称	项目总数
	数据访问服务	5
	数据分类与溯源	5
	错误管理	2
	全局事件服务	4
	网格信息服务	4
	全局请求	0
	监测	14
	服务质量的可靠性	1
	资源访问服务	20
	安全与隐私	10

3.3.2 软件组件

软件是网格项目至关重要的组成部分。美国的各种机构和组织开发了大量的网格软件组件，而且其中大部分都是开放源代码的，理解这些组件的能力、优缺点以及目前所能够解决的问题的类型将能够为新的网格建设解决很多实际的问题。另外，每一个新的网格项目也都有各自唯一需求，需要用一组特定的软件组合来实现。

仅从一次应用项目中得出的通用组件并不是真正意义上的通用，只有被使用在多个项目中，并且能够保持一定稳定性和精良性，才是对通用最好的定义。一些通用的解决方案，如 Globus Toolkit 的安全、信息服务，作业管理，数据传输组件等，MyProxy 作为稳定的存储库，SDSC 的 Ganglia 聚类监测工具，以及 DataGrid 项目的 VOMS 角色管理服务等等。对于一组通用解决方案，组合使用它们可以产生很多成果，根据具体的应用和系统，采取不同的方法整合这些组件能够产生高层次的解决方案，显示出更强大的功能，如 LIGO 中的 Lightweight Data Replicator 系统。下面从 6 个方面对美国开发和使用的组件进行总结。

1. 网络安全

管理网格应用程序的认证在理论上是件很容易的事情，但在实践过程中却极具挑战性。网格研究人员已经开发了认证管理工具和服务以便为用户和服务生成凭证，从而使用户可以登陆使用网格，并且在系统中需要的时候得到用户的网格凭证。这些认证管理工具和服务不仅仅可用于确认用户和服务的身份，也可以通过服务对基本的网络安全机制实施访问控制。目前，已有研究人员开发出授权和访问控制工具

以在系统的授权范围内存储和提供访问，并通过创建中央数据存储来提供分散控制机制。以下列出相应的软件组件。

- ① 基本安全机制：Pre-Web Services Authentication and Authorization、Web Services Authentication and Authorization;
- ② 凭证生成组件：Globus Certificate Service、Simple CA;
- ③ 凭证管理组件：MyProxy、KX.509 and KCA、PKINIT;
- ④ 访问控制和授权组件：Shibboleth、Community Authorization Service (CAS)、VOMS。

2. 网络监测与发现

网络系统和应用通常应该是保持稳定，但框架的特定部分（如计算单元、存储系统、设备和遥感站等）可能掉线或被动态地添加到系统中，因此系统或应用需要很好的使用框架元素的这种动态性，以提供稳定持续的服务。这就需要检测和发现框架中的元素和服务。OGSA 为发布、定位和订阅信息提供了一个核心的框架与实现，而且网络研究人员也已经开发出几种相应的监测系统：

- ① 基本监测与发现机制：WS Core Monitoring Features、Globus Index Service;
- ② 特定的监测与发现组件：Globus Trigger Service、Ganglia Cluster Toolkit、Inca、MonALISA。

3. 网络计算

分布计算是网络计算研究较早的领域之一，因此存在很多工具。但在人们在一个计算资源上执行作业或使用服务时，仍然存在很多问题，如账目管理、沙箱技术（sandboxing）、持续的运行环境等方面存在的问题，因此还需要继续研究。目前已有的解决方案有：

- ① 基本的计算机制：GRAM、GRAM Scheduler Plugins;
- ② 分布编程框架：NetSolve/GridSolve、Ninf-G、MPICH-G2、Condor-G、DAGman;
- ③ 元调度组件：Platform CSF、GridWay;
- ④ 虚拟数据组件：Chimera。

4. 网络数据管理

传感器网络和卫星图像系统的部署以及这些图像系统的解决方案导致大量原始数据的获取。这些数据，在逐渐增强的计算能力的支持下，产生了大量的数据分析结果，因此迫切需要大容量的数据存储和管理系统。与此同时，仅仅存储和检索数据也是不够的，数据必须在合作项目中能够对每个参与者来说都是可以使用的，并且在不同地区都能够以最快的速度访问，使用描述信息进行分类以使其便于检索，并且让网络上正在运行的计算作业能够访问。为建设这样一个高容量的存储系统与

网络，网络研究人员们生产了一系列的组件以在网络上管理数据和使用。

- ① 基本数据管理机制：GridFTP、OGSA-DAI、Metadata Catalog Service (MCS);
- ② 移动和传输数据组件：globus-url-copy、Reliable File Transfer (RFT) Service、UberFTP、GSI-SCP/SFTP;
- ③ 最优化数据访问组件：Replica Location Service (RLS)、NeST、DataCutter。

5. 网络合作

网络已经被开发到使得支持合作的应用和系统变得容易。从某种意义上说，网络中所有组件都是为了方便合作而设计和使用的。但是，大多数网络组件是用来解决系统问题的，如安全问题、数据访问问题、计算问题等，而另外还有一些组件则更加关注用户层面的问题，如用户界面、人与人的通信（聊天、电子邮件、公告板、远程会议等），以及团队建设（如文件共享、通知等），这些组件包括：

- ① 信息门户建设组件：Sakai、pen Grid Collaboration Environment (OGCE);
- ② 功能强大的虚拟会议空间：Access Grid。

6. 软件打包和分配工具

网格系统已经有了很多组件，但多数组件只具备较为狭窄的使用功能，而给最终用户建立一个有用的网格应用或系统需要将这些组件整合到一起，并增加其他的面向特定应用的组件，并在一系列物理资源上以特定的方法部署和配置所有这些以得到需要的结果。这种整合组件的能力，对于任何一个网格项目的成败来说都是至关重要的。在早期的网格项目中，系统开发人员一般都从一个预先存在的网格组件分配开始，比如一个“整合分配单”。随着网格项目向着“产品”阶段的转变，系统操作员为节省投入一般只要求最少的组件配置，在这种情况下，以下的组件打包和分配工具就显得格外有用：

- ① 分配和打包工具：PACMAN、Grid Packaging Tools (GPT);
- ② 已经整合好的工具包：NSF Middleware Initiative (NMI)、Rocks、Virtual Data Toolkit (VDT)。

3.4 美国e-Science当前进展

本节简要介绍美国 e-Science 项目在资源层、连接层、网络中间件方面的进展情况。

在资源层上，在科学研究的资源方面，美国无疑是世界上最为丰富的国家之一，从数量众多的高校到实验室、从观测站到试验设备、更有大量的计算资源。

在连接层上，美国国家骨干网 10Gb 的带宽为 e-Science 的建设和发展提供了网络连接速度和质量上的保障，目前的各种网络建设项目，如 UltraLight 等，则会在更

大程度上提高美国科研的硬件保障水平，再加上如 NSF Office of Cyberinfrastructure 这样的机构所资助项目的研究，使得利用这些网络资源更加有效。

在网格中间件的建设方面，美国已在全国范围内建立了大批的网格，其中包括基础型网格设施，如 TeraGrid，DOE Science Grid 等，更有为数众多的应用型网格，如 Access Grid，IPG（Information Power Grid），GriPhyN（Grid Physics Network），PPDG（The Particle Physics Data Grid），ATLAS（Alphanumeric Grid），GEON（GEOsciences Network），ESG（Earth System Grid），BIRN（Biomedical Informatics Research Network）等等，以及参与国际合作的网格系统，如 iVDGL（International Virtual Data Grid Laboratory），STAR grid，PRAGMA（Pacific Rim Applications and Grid Middleware Assembly），NVO AP Grid（Asia-Pacific Grid）等，可以说面向科学研究的各个领域，为大规模的科研合作，大量的科学计算，大规模的数据存储、管理、移动、分析，组织间通信等都提供了前所未有的技术支持，并极大的推动了科研的发展。而且，已建立的网格系统也将全国的高校、实验室、计算中心等科研与科研辅助机构连接到了一起，形成了全国性的研究网络，截止于 2004 年，仅 Grid 3 一个项目就连接了 35 个站点、3500 台 CPU、4 个国家实验室和数量众多的高校，而基于这一国家型的基础网格，又运行着 HEP（High Energy Physics）、LIGO（Laser Interferometer Gravitational-wave Observatory）、SDSS（Sloan Digital Sky Survey）、GENO 等网格项目，后来随着 iVDGL，GriPhyN，PPDG，US LHC（Large Hadron Collider）的加入，组成了 DOE Science Grid。

类似 Grid 3 这样的国家网格基础设施，在美国还有很多，如 TeraGrid 等，并且在 OGSA 的支持下，这些大型网格正在走向融合，实现了真正的资源共享、服务共享。

对于美国 e-Science 研究的项目，前文已经从资助、承担、项目应用和项目组成等几个角度进行了较为详细的分析和介绍，这里还需要对项目的研究成果做一个简单的介绍。在调研的美国 e-Science 项目中，项目的成果形式包括：网格系统、软件组件、工具包、应用程序、服务、资源、框架、标注、协议、硬件系统、门户网站、研究报告和试验床（testbed），具体分布情况如图 3-10 所示。其中，共包括网格系统 32 个，由此也可以看出美国网格研究和建设的繁荣景象，另外在调研项目中有 38 个项目涉及到了软件组件的开发，15 个项目涉及工具包的研究，13 个项目涉及到技术框架的探索，这一方面说明人们对 e-Science 技术研究的重视程度，而更深层次的意义在于，尽管技术框架还处于发展的过程当中，但更多的研究重点已经不在于框架的讨论中，而是放到了对具体组件的实现和工具包的组装方面，这可以从某个角度说明美国的 e-Science 技术研究，或者说是网格研究已经进入了一个比较稳定的发展阶段。当然近 1/3 应用型项目这一数字也可以从另一个角度说明同样的问题。此外，在所有软件产品的项目中，80% 的项目的软件产品是开放源代码的。

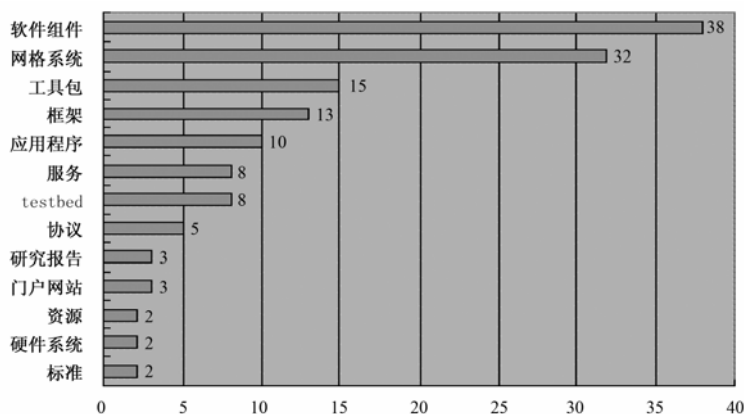


图 3-10 美国 e-Science 项目成果类型分布

3.5 小 结

尽管此次调研没有网罗到美国所有 e-Science 研究方面的项目，但由于采用了一定的检索方法和标准，所以通过对调研项目的分析和总结，也可以说能够归纳出美国在 e-Science 领域的研究以下特点：

- 没有统一的管理机构和计划，美国 e-Science 的研究完全是在科研需求的基础上自由发展起来的；
- 项目主要由国家机构资助，而且以 NSF 和 DOE 为主；
- 科研机构是主要的研究力量，其中又以国家实验室为最，高校是主要的参与研究机构；
- 研究内容更多的关注于对 e-Science 的技术实现方面，在应用领域则主要集中于地球科学和物理学研究；
- 在研究过程中，与欧洲的合作较为密切，其中与英国的合作最多；
- 在技术研发中，更关注软件组件的开发和组装问题；
- 基于美国先进的网络基础设施和丰富的科研资源，已经建立了大规模的全国性以及世界范围内的科研合作体系。

总体来说，美国的 e-Science 研究和建设正在新时代大规模科学研究所带来跨地域科研合作、海量数据存储、传输、处理和分析、仪器设备等各种研究资源贡献、复杂科学计算和其他如科学模拟和可视化等的挑战下，以及科研工作者的创造力和坚持不懈的努力下蓬勃的发展着。

第 4 章 欧洲 e-Science 发展现状分析

4.1 欧洲 e-Science 历史背景

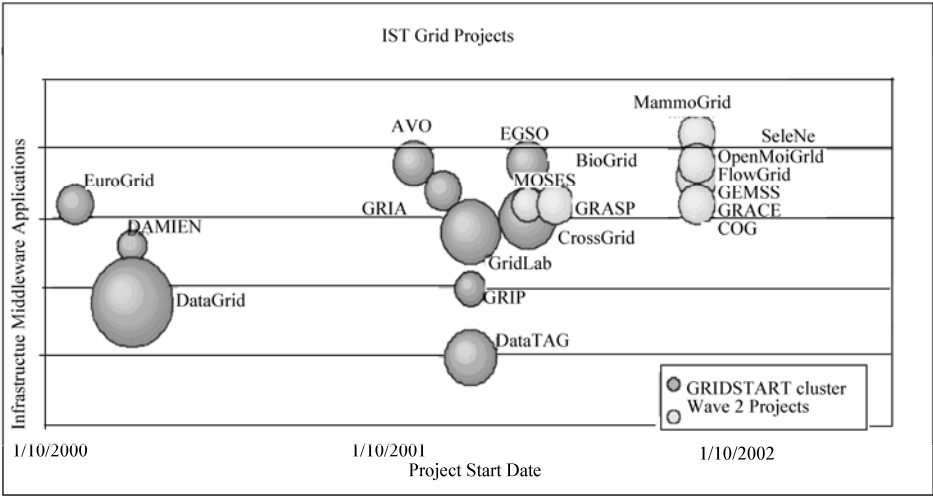
e-Science 概念最早是由英国科学技术办公室 (OST) 在其 2000 年宣布资助的英国 e-Science 计划中提出。该计划的负责人暨英国研究委员会主任 John Taylor 给 e-Science 下的定义是:“e-Science 是指关键科学领域中的全球化合作以及使这种全球化合作得以实现的下一代基础设施。”

e-Science 的基本内涵包含三个层次的内容,即基础设施、核心技术和应用。按照 e-Science 的这个定义和基本内涵来追溯欧洲地区的 e-Science 发展历史,欧洲地区尤其是欧盟国家在 20 世纪 90 年代已经开始启动项目研究 e-Science 的基础设施和核心关键技术。在欧洲地区的 e-Science 发展过程中,网格研究始终贯穿其中,而网格作为下一代的研究基础设施,直接影响着各国 e-Science 的发展水平。

欧洲地区 e-Science 的发展沿着三条主线进行,分别是欧盟的 e-Science 计划和项目、欧洲其他地区性的 e-Science 计划和项目以及欧洲地区各国的 e-Science 计划和项目。由于英国 e-Science 计划和项目相对比较多,因此本节关于欧洲地区的 e-Science 进展将不包括英国地区。

1. 欧盟 e-Science 的历史

欧盟最早关于 e-Science 的研究计划始于 2000 年年初。欧盟委员会在第五研究框架计划(The Fifth Framework Programme,FP5)中第一次开始资助网格研究项目。在 FP5 下的网格研究项目集中在网格技术开发和网格应用试验。欧盟在 FP5 下总共资助了 24 个网格研究项目,总资助额达 5800 万欧元。欧盟在 FP5 下的网格研究项目分两个时间段启动,第一阶段从 2000 年年底至 2001 年年底,总共启动了 10 个网格研究项目;第二阶段从 2002 年年初至 2002 年年底,总共启动了 10 个项目(图 4-1)。在 2002—2006 年的欧盟第六研究框架计划(The Sixth Framework Programme ,FP6)中,欧盟委员会加大了对网格研究项目的资助额度,欧盟在 FP6 下的网格研究项目资助预算将是超过 FP5 下资助总额的两倍,可见网格研究在欧盟中的战略重要性。FP6 下的网格研究将同时注重技术推动和应用拉动,并将强调开发开放标准和利用开放资源。从 2004 年 4 月份起,FP6 启动了第一批 12 个网格研究项目,总资助额达 5200 万欧元。欧盟认为网格将使整个欧洲地区的研究者更好地分享知识和各种资源,网格将真正支撑欧洲研究领域(ERA),网格将会整合整个欧洲的世界一流的科技资本以增强欧洲的竞争力和生活质量。因此网格将是实现欧盟 2000 年设定的“里斯本战略”目标的关键技术。



Priorities of European Garid Projects(the size of the“bubbles” denotes the funding level of the corresponding projeece)

图 4-1 欧盟 FP5 下网格项目启动的时间分布

2. 欧洲其他地区e-Science的历史

除了欧盟之外，欧洲其他地区在 e-Science 发展中也取得了一定进展。2001 年 5 月，北欧地区四国挪威、丹麦、瑞典和芬兰共同启动了名为 NorduGrid 的网格计划。该计划旨在为北欧地区提供一个健壮的、可升级的和便携的全球计算和数据网格系统的全面解决方案。2002 年东南欧地区希腊、匈牙利等十个国家共同启动了名为 SEE-GRID 的网格研究项目。该项目旨在为东南欧地区的国家参与泛欧洲和世界范围内的网格计划铺平道路。此外，欧洲波罗的海地区的瑞典、波兰等五国也启动了名为 Baltic Grid 的网格研究项目。该项目旨在使波罗的海地区的网格研究融入整个欧洲网格基础设施研究，并希望波罗的海地区国家的网格技术达到欧盟国家的一个可比水平。

在欧盟和欧洲地区性的 e-Science 项目启动之前，欧洲地区各国已经开始启动一些国家性 e-Science 研究项目。

3. 德国e-Science的历史

德国是欧盟国家中较早开展 e-Science 研究的国家之一。德国早在 20 世纪 90 年代早期就已经开始研究网格系统。1997 年 8 月启动的 UNICORE 项目是德国早期比较有代表性的网格项目，该项目的目标是提供一种科学与工程网格，该网格可以将资源和超级计算机联结起来，并使得这些资源和计算机通过因特网可以被研究人员使用。UNICORE 创建了访问超级计算机的无缝高性能计算（HPC）入口。目前，

UNICORE 系统的第一代产品已经在 2000 年年底投入到德国 HPC 中心使用。2000 年 1 月, 德国启动了 UNICORE Plus 项目。该项目将在 UNICOERE 的基础上进一步为科学家和工程师开发一个在因特网任何地方都能访问超级计算中心的网格基础设施和计算入口。但是 UNICORE 和 UNICORE Plus 只是解决了访问超级计算资源的问题, 德国仍然没有完全支持 e-Science 的基础设施。因此, 在 2002 年 12 月 UNICORE Plus 的项目评估会上, 德国提出了国家 e-Science 计划。2004 年 9 月德国联邦教育与研究部 (BMBF) 在国家 e-Science 计划下资助一个名为 eSciDoc 的 e-Science 项目, 该项目由德国著名的研究机构 FIZ Karlsruhe 和马普学会共同承担, 总资助额为 610 万欧元。该项目的目的是为马普学会 80 个研究所的科学家开发一个基于网络的科研工作合作平台。2005 年 9 月德国启动了国家 e-Science 框架计划——D-Grid 计划。该计划是目前德国最大的 e-Science 计划, 它将为德国的网格技术和 e-Science 的可持续发展建立基础。该计划的目标是协调和加强德国国内现有的网格计算活动。D-Grid 计划将持续到 2008 年, 它的第一个目标是开发一个连接现有网格计划的国家 e-Science 基础设施。

4. 法国e-Science的历史

法国也是欧洲比较早进行 e-Science 研究的国家, 早在 1995 年法国就首次尝试不同计算机族之间的互联。1996 年法国里昂计算机科学实验室 (LIP) 邀请 Globus 的创始人 Ian Foster 在欧洲并行计算大会 (Euro-Par) 上介绍网格计算的研究成果。法国从那时起开始关注并跟踪研究 e-Science。1998 年法国启动了国家网络互联和电信网络 (RNRT) 计划, 该计划由法国工业和电信部联合资助。在 1999—2002 年四年间该计划总共资助了 200 多个项目, 其中来自公共资助额达 1 亿 8500 万欧元, 而私有资助额达 2 亿 1100 万欧元。2000 年法国研究部和工业部共同资助发起了国家软件工程网络 (RNTL) 计划。在 2000—2002 年三年间共有 120 个项目被资助, 总资助额达 1 亿 9100 万欧元。RNRT 和 RNTL 这两个计划在法国网格研究中有重要作用, 它们为法国学术界和工业界提供了网格研究的平台。2001 年启动的 ACI GRID 计划是法国 e-Science 发展中的重要一步, 该计划由法国研究部国家科学基金 (FNS) 单独资助, 该计划每年资助 15 个项目, 平均每年的资助预算是 300 万欧元。ACI GRID 计划在支持法国学术团体开展网格技术基础研究上扮演着重要角色。在 2003 年启动的 GRID 5000 项目就是在 ACI GRID 计划资助下产生的项目。GRID 5000 项目的目的是将全国范围内八个地区的 5000 个处理器通过高速网络互联在一起, 为法国的网格研究建立一个大型平台。

5. 意大利e-Science的历史

意大利在 20 世纪 90 年代后期开始 e-Science 研究。1999 年国家核物理研究所 (INFN) 发起的 INFNGrid 项目是意大利的第一个网格项目。意大利的 e-Science 研

究沿着两条主线进行。一条是在高性能计算和元计算主题下产生的一系列研究项目，主要由意大利国家研究委员会（CNR）牵头，并与意大利空间局（ASI）、INFN（Istituto Nazionale di Fisica Nucleare）、ENEA（National Agency for New Technologies, Energy and the Environment）等科研机构及高校及工业界合作。另一条是开发 e-Science 的网格基础设施，主要由 INFN 牵头，并与 CERN、高校以及工业界等合作。这两条主线现在已经被整合进意大利国家网格研究项目 GRID.IT。该项目由意大利教育大学和研究所（MIUR）在 2001 年 8 月发起，并在 2002 年 9 月被正式批准。该项目总资助额为 810 万欧元。该项目旨在定义、实施和应用面向虚拟组织和网格计算的网格计算平台的创新解决方案。

6. 荷兰e-Science的历史

荷兰是欧洲较早从事 e-Science 研究的国家。2001 年阿姆斯特丹科技中心（WTCW）在荷兰组织了全球网格论坛（GGF）第一次会议。荷兰单独成立了国家网络组织 SURF，该组织最近获得资助在荷兰建立第六代网络基础设施 SurfNet6。此外，荷兰高级计算和成像研究生院（ASCI）最近也启动了并行和分布式系统的超级计算项目 DAS。DutchGrid 是荷兰网格计算和网格技术的研究平台，它的目的是协调荷兰国内各种网格活动的部署以及为网格技术的经验交流提供一个论坛。WTCW 和阿姆斯特丹大学等 4 个科研机构共同发起了面向 e-Science 的虚拟实验室（VL-e）项目。该项目的目标是创建一个 e-Science 环境以促进 e-Science 研究。

此外，欧洲的其他国家也先后开展了一些 e-Science 研究项目，包括瑞典的 SweGrid；瑞士的 SwissBioGrid；西班牙的 IrisGrid；希腊的 HellasGrid；匈牙利的 ClusterGrid、DemoGrid、SuperGrid、JGrid；捷克的 MediGrid；波兰的 PROGRESS、PIONIER、SGIGrid；奥地利的 eScience Austria、ASKALON、Austrian Grid；爱尔兰的 Grid-Ireland 和 CosmoGrid 等。

4.2 欧洲e-Science承担主体分析

本节对欧盟在第五框架计划（1998—2002 年）和第六框架计划（2002—2006 年）下资助的共计 40 个项目（见附录 B）的参与机构性质、参与机构数目、机构参与项目数目、参与国家数目、各国参与项目数目、项目负责人所在国家、项目资助额度、项目启动时间和项目资助期限共九个方面进行了统计分析。

1. 参与机构性质分析

欧盟的这 40 个 e-Science 项目总共有 520 个机构参与。从参与机构的性质上看，其中最多的是科研机构，共有 199 个，占全部参与机构的 38%；其次是高校院系，

共有 182 个，占总数的 35%；企业的参与数目为 139 个，占全部参与机构的 27%。总的来说，欧盟 e-Science 项目的三个承担主体比较平衡（图 4-2）。

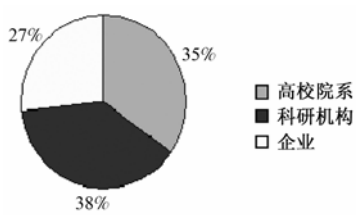


图 4-2 e-Science 项目参与机构数目按机构性质分布

2. 参与机构数目分析

在这 40 个项目中，一个项目至少有 5 个研究机构同时参与，而最多一个项目有 71 个机构同时参与。在所有项目中，参与机构数目在 5~10 个之间的项目数目最多，共有 25 个，占全部项目的 62%；其次是参与机构数目在 21~71 个之间的项目，共有 8 个，占全部项目的 20%；最后是参与机构数目在 11~20 之间的项目，共有 7 个，占全部项目的 18%（图 4-3）。可见，欧盟资助的单个 e-Science 项目的参与机构数目大多数都在 10 个以内。

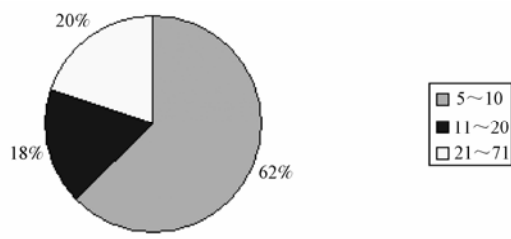


图 4-3 e-Science 项目按参与机构数目分布

3. 机构参与项目数分析

本节对欧盟第五框架计划的 20 个项目和第六框架计划的 14 个项目的参与机构进行了统计，总共有 234 个机构。在所有这些机构中，按参与项目数统计排在前五位的机构分别是德国斯图加特高性能计算中心（HLRS）、德国 Forschungszentrum Juelich 研究中心（FZJ）、英国曼切斯特维多利亚大学（UMAN）、德国柏林 Zuse 研究所（ZIB）和德国 Fraunhofer 研究所。在前五位中有四个机构来自德国，可见德国在欧盟内 e-Science 研究中有着重要作用。此外，图 4-4 给出了按参与项目数排在前 16 位的机构名称。可以说，这些机构是欧盟 e-Science 研究的主要力量。

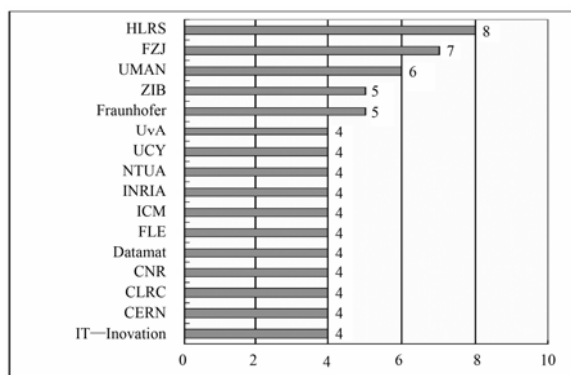


图 4-4 按参与 e-Science 项目数排在前 16 位的研究机构

4. 参与国家数目分析

欧盟资助的 e-Science 项目在参与国家数目这个指标上比较有特色。由于欧盟目前总共有 25 个成员国，因此欧盟资助的项目一般都是由欧盟内部成员国共同参与。在这所有 40 个项目中，一个项目至少有 2 个国家参与，而最多的一个项目则有 27 个国家共同参与。在所有项目中，参与国家数目在 2~5 个之间的项目数最多，总共有 22 个，占全部项目的 54%；其次是参与国家数目在 6~10 个之间的项目，共有 15 个，占全部项目的 38%；而参与国家数目在 11~27 个之间的只有 3 个项目，占全部项目的 8%（图 4-5）。此外，在这 40 个项目中的 7 个项目有欧盟以外的其他国家参与。

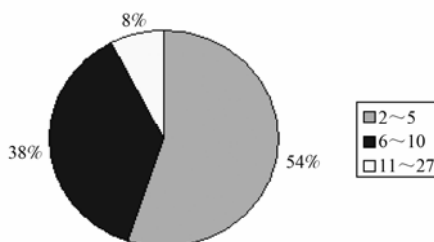


图 4-5 e-Science 项目按参与国家数目分布

5. 各国参与项目数目分析

从欧盟国家参与第五框架计划和第六框架计划的项目数目上分析：（1）目前欧盟 25 个国家中除了拉脱维亚、马耳他和卢森堡 3 个国家外，其他 22 个欧盟国家至少参与了其中一个项目，可见 e-Science 研究在欧盟内部具有广泛性。（2）欧盟各国 e-Science 研究力度存在差异，按参与项目数目排在前 5 位的国家分别是德国、英国、意大利、法国和西班牙。这些国家在 e-Science 研究中相当活跃，其中德国在所有 40 个项目中参与了 31 个，参与度达 78%。（3）欧盟在 e-Science 研究上注重与欧盟以

外的欧洲国家和世界上其他国家进行合作，这些国家包括挪威、瑞士、美国、俄罗斯和中国等（图 4-6）。

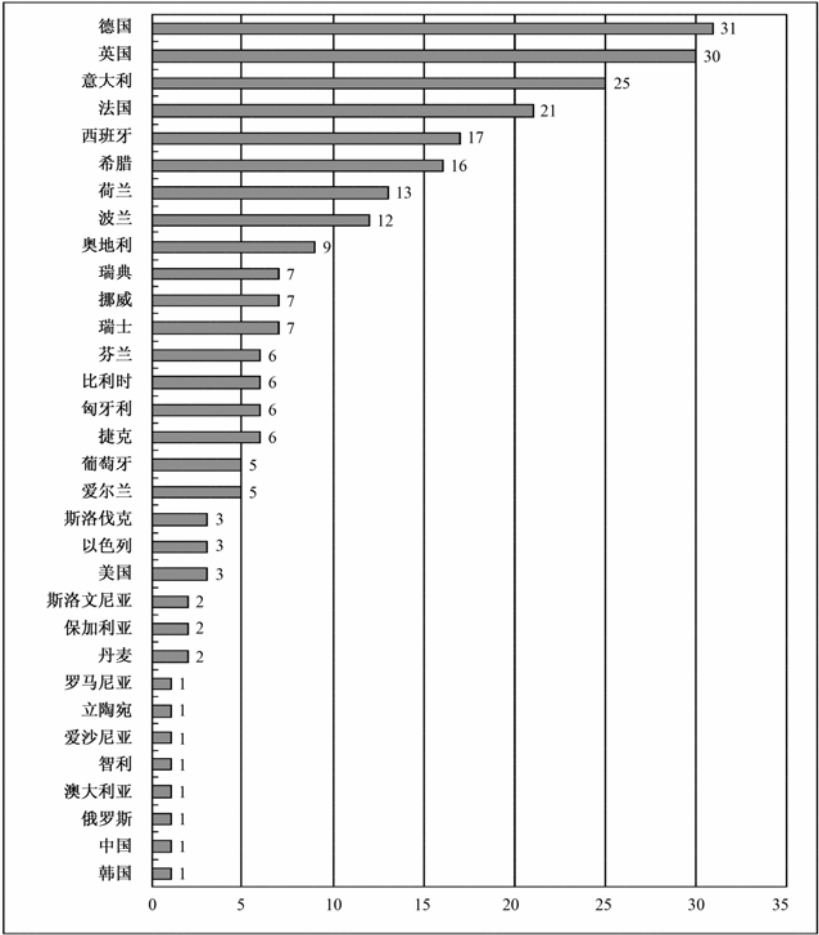


图 4-6 每个国家参与的 e-Science 项目统计

6. 项目负责人所在国家分析

从对项目的负责人所在国家这个指标看，欧盟 25 个国家中只有 8 个国家作为项目的负责人，另外还有一个欧洲原子能研究组织（CERN）。在所有 40 个项目中，英国和德国作为项目负责人的项目数最多，都是 9 个，分别占全部项目的 22%；其次是西班牙，作为项目负责人的项目数有 5 个，占全部项目的 12%；法国和意大利作为项目负责人的项目数都是 4 个；CERN 负责 3 个项目；波兰作为负责人的项目有 2 个；希腊和斯洛文尼亚各负责一个项目；另外还有 2 个项目的负责国家未知（图 4-7）。

7. 项目资助额度分析

从欧盟 FP5 和 FP6 的资助额度上看，欧盟资助额在 100 万～500 万欧元之间的项目最多，总共有 23 个，占全部项目的 57%；其次是 100 万欧元以下的项目，共有 5 个，占全部项目的 12%；资助额在 500 万～1000 万欧元之间的项目有 5 个，占全部项目的 13%；资助额在 1000 万欧元以上的项目有 3 个，占全部项目的 8%；另外还有 4 个项目的资助额未知（图 4-8）。

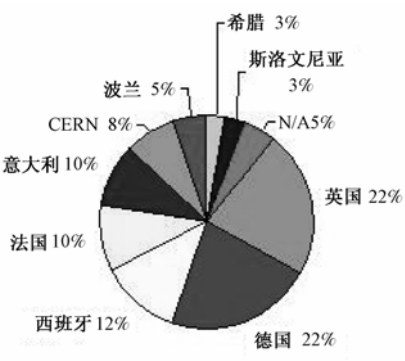


图 4-7 e-Science 项目按项目负责人所在国家分布

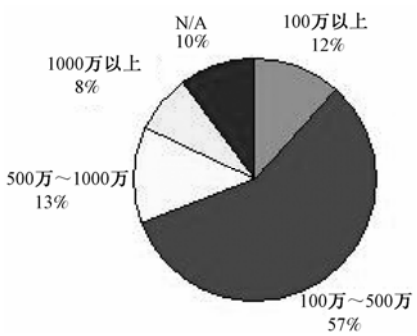


图 4-8 e-Science 项目按资助额度分布（单位：欧元）

8. 项目启动时间分析

欧盟第五框架计划其实从 1998 年已经开始启动，但是 FP5 资助的 e-Science 项目到 2000 年才开始启动。FP5 第一个网格项目在 2000 年年底启动，然后在 2001 年又启动了 4 个项目，FP5 在 2002 年启动的网格项目达到了一个高峰，整个 2002 年总共启动了 19 个项目。随后在欧盟第六框架中，网格项目的重要性进一步被强调，欧盟在 FP6 下的资助额将超过 FP5 的两倍，在 2004 年 FP6 下总共启动了 15 个项目，在 2005 年目前也已经启动了 1 个项目（图 4-9）。

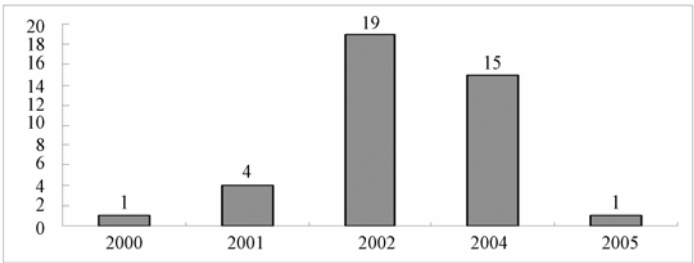


图 4-9 e-Science 项目按项目开始年份分布

9. 项目资助年限分析

欧盟资助的项目期限来看，这些项目的资助年限都超过 1 年，最长的项目资助年限是 4 年。资助期限在 1 年以上到 2 年的项目有 4 个；资助年限为 2 年的项目有 7 个；资助年限在 2 年以上到 3 年的项目数量最多，总共有 12 个，占全部项目的 30%；资助年限为 3 年的项目有 10 个，数量上居第二位，占全部项目的 25%；资助年限在 3 年以上到 4 年的项目有 3 个；资助年限为 4 年的项目有 3 个；另外还有 1 个项目的资助年限未知（图 4-10）。

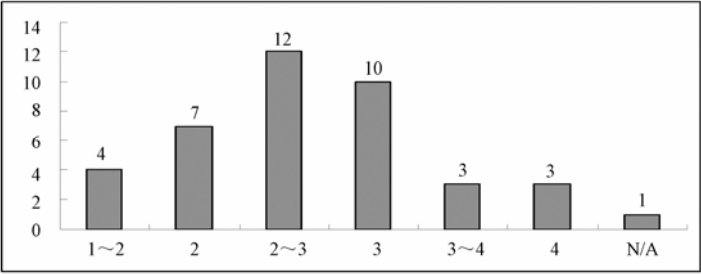


图 4-10 e-Science 项目按资助年限分布

4.3 欧洲e-Science技术框架与技术模块

4.3.1 网络技术层次分析

欧盟的网格研究将网格技术划分为 4 个层次，分别为网格应用和入口（Applications and Portals）、应用开发环境和工具（Application Development Environments and Tools）、中间件（Middleware）和基础设施（Fabric and Connectivity）（图 4-11）。

从网格技术层次的角度来分析欧盟第五框架计划和第六框架计划下的网格项目（详见附录 C），可以发现欧盟网格研究的技术演变。（1）欧盟第五框架计划第一阶段资助的网格项目集中在网格技术的较低层次，其中大部分项目都集中在网格基础设施和网格中间件两个层次，而且这一阶段的资助特点是资助项目数量少但资助金额大；（2）第五框架计划第二阶段资助的网格项目则集中在网格技术的较高层次，主要是开发特定领域内的网格应用，这一阶段欧盟资助的特点是资助项目数量多但资助金额小；（3）第六框架计划资助的网格项目继续集中在网格技术的较高层次，并已经开始从实验性网格到生产性网格（Production Grid）转移。可见，欧盟的网格研究已经从最初的网格基础设施研究过渡到网格应用领域研究，这种过渡也说明欧盟早期在网格基础设施方面的研究已经为网格应用研究提供一个很好的基础。此外，

欧盟早期的网格研究主要集中在分布式超级计算、按需计算和数据密集型应用方面，而欧盟最近的网格研究则更侧重协同合作应用研究、面向知识发现的应用研究。在欧盟第五框架计划下的 20 个网格项目，按照网格技术的 4 个层次来划分，其中研究网格应用和入口层次的网格项目最多，总共有 20 个；其次是研究中间件这个层次的网格项目，共有 18 个项目；研究网格应用开发环境和工具的项目有 11 个；而 20 个项目中涉及网格基础设施研究的最少，只有 9 个项目（图 4-12）。此外，FP5 大型的网格项目的研究内容覆盖了所有网格技术的 4 个层次，比如 DataGrid、GridLab、CrossGrid，而一些小型的项目只研究网格技术中的特定层次，比如 BioGrid 项目只研究网格技术应用和入口这一层次，GRACE 项目只研究中间件和应用与入口两个层次。从时间维度看，基本上 FP5 第二阶段的项目都属于集中研究网格技术高层或特定应用领域的小型项目。

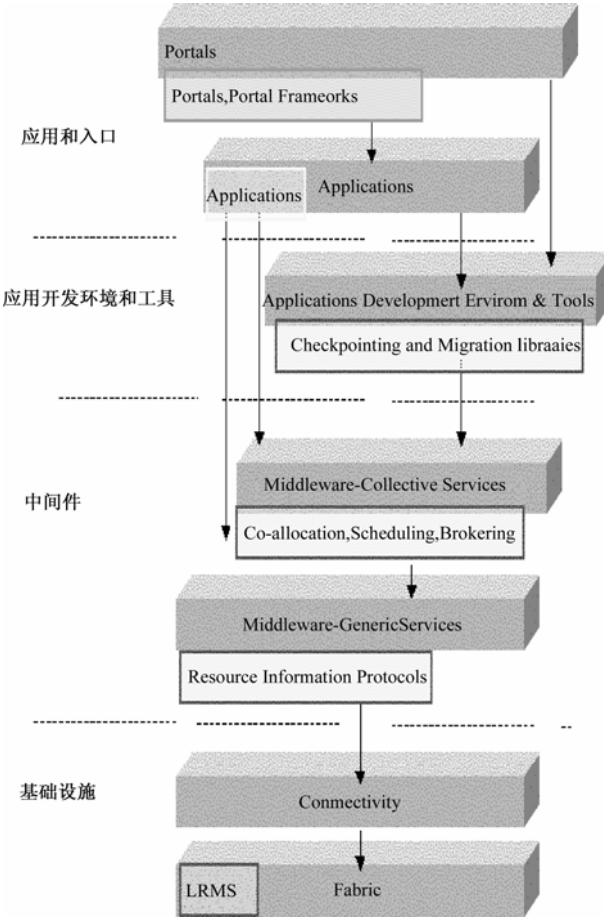


图 4-11 网格技术的四个层次

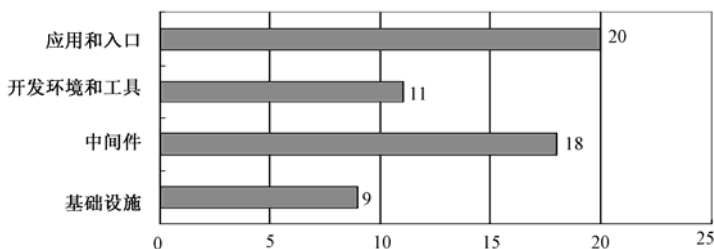


图 4-12 FP5 项目按网络技术层次分布

4.3.2 网格技术开发领域分析

欧盟的网格研究按网格技术开发领域（Development Area）划分为 12 个网格开发领域，分别是：（1）网格入口和移动访问；（2）应用；（3）应用环境和工具；（4）发布；（5）测试平台和网络；（6）数据管理；（7）信息服务；（8）登录管理；（9）监测和绩效；（10）会计；（11）资源管理和日程安排；（12）安全。

对第五框架计划下 20 个网格项目按照这 12 个网格技术开发领域进行分析后发现，FP5 计划研究应用领域的网格项目最多，总共有 17 个。12 个开发领域中排在前五位的开发领域依次为网格应用、发布、测试平台和网络、数据管理、应用环境和工具、监测和绩效。在前 5 个领域，每个领域都至少有 10 个以上的项目进行研究。而这 12 个领域中研究较少的是网格登录管理和会计，分别只有 3 个和 7 个项目对其进行研究（图 4-13）。

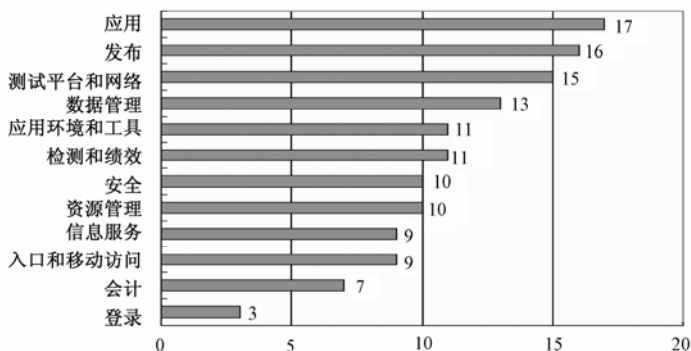


图 4-13 FP5 项目按网格开发领域分布

4.3.3 网格技术应用类型分析

欧盟将网格研究按其应用类型划分为 5 大类，分别是：（1）分布式超级计算（Distributed Supercomputing）；（2）高性能计算（High Throughput Computing）；（3）按需计算（On-demand Computing）；（4）数据密集型应用（Data Intensive

Applications)；(5) 协作应用 (Collaborative Applications)。在欧盟第五框架计划的 20 个项目中，研究按需计算的有 9 个项目；研究数据密集型应用的有 8 个项目；此外，分别有 7 个项目研究分布式超级计算和协作应用；另外还有 2 个项目研究高性能计算（图 4-14）。

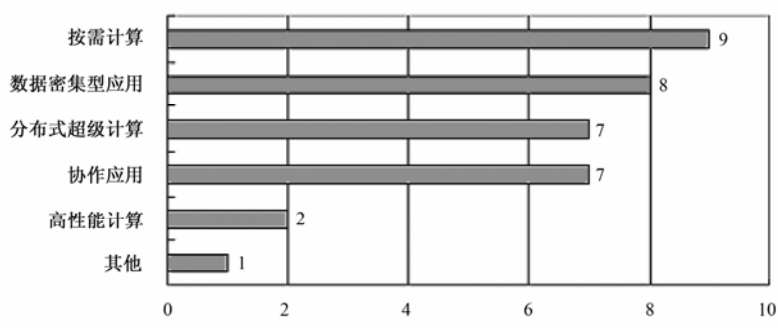


图 4-14 FP5 项目按网格应用类型分布

4.4 欧洲e-Science当前进展

当前欧盟的 e-Science 研究已经取得了重要成就。欧盟第五框架计划资助的网格研究项目已经全部结束，而第六框架计划下的网格项目也已经在 2004 年全面启动，e-Science 必将在欧盟新一轮框架计划中进一步体现其战略重要性。在日前欧盟公布的一份评估第五框架计划下网格项目的报告中，GRIDSTART 委员会将第五框架下网格研究取得的成就概括为下面六点：

- ① 创造了一个强大的网格研究团体；
- ② 在网格中间件开发和标准化方面增强了欧洲在全球的地位；
- ③ 在面向特定应用需求的垂直中间件领域确立了领导地位；
- ④ 在为工业和商业用途的成熟的网格技术研究方面迈出了第一步；
- ⑤ 网格在 e-Science 应用试点中进一步被证明有效；
- ⑥ 为服务提供商识别了正在出现的机会，并认识到网格技术要被更广泛地采用，让其更加商业化是至关重要的。

4.4.1 项目技术成就分析

欧盟第五框架计划的网格项目在技术上的成就主要体现在七个方面：(1) 网格标准化；(2) 中间件开发；(3) 商业化应用；(4) 科学应用；(5) 互操作；(6) 网格开发工具；(7) Web 服务。GRIDSTART 委员会在总结第五框架计划下网格项目成就时具体给出了 16 个项目在这七个技术成就上的统计（表 4-1）。

表 4-1 第五框架计划网格项目的技术成就

技术成就/ 项目名称	网格 标准化	中间 件开发	商业 化应用	科学应用	互操作	网格 开发工具	Web 服务
BioGrid			•			•	•
CrossGrid		•		•	•	•	
Damien		•	•			•	
DataGrid	•	•		•		•	
EGSO				•	•	•	•
EuroGrid	•	•	•	•			
FlowGrid		•	•			•	
GEMSS		•	•				•
GRACE	•	•	•	•			
GRASP		•	•			•	•
GRIA		•	•			•	•
GridLab		•		•		•	
GRIP	•	•			•		
MammoGrid			•	•			•
OpenMolGrid			•	•		•	
WebSI		•	•				•

4.4.2 项目开发出的组件分析

欧盟第五框架计划的网格项目在 12 个网格开发领域总共开发出了 43 个组件（详见附录 D），其中网格监测和绩效领域开发的组件数目最多，总共开发出了 11 个组件；其次是网格入口领域开发出了 6 个组件；网格应用、应用环境和工具以及资源管理和日程安排这三个领域分别开发出了 5 个组件；信息服务、安全和数据管理三个领域则分别开发出了 3 个组件；最后在登录管理和测试平台两个领域分别开发出了 1 个组件(图 4-15)。

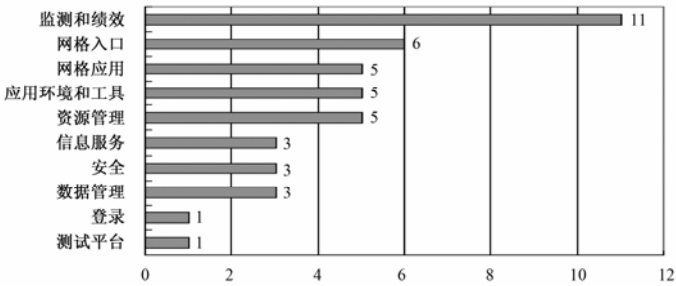


图 4-15 FP5 项目开发的组件数按开发领域统计

4.4.3 项目应用成就分析

欧盟第五框架计划下的网格项目除了在技术上取得了重要成就外，在各个应用领域也取得了广泛的成就。**FP5** 项目研究涉及的应用领域包括生物学和医学、地球和环境科学、工程学、天体物理学、物理学、工商业、教育、语义网和多媒体等。其中研究生物学和医学的项目最多，总共有 10 个，其次是地球与环境科学和工程学各有 5 个项目（图 4-16）。此外，从时间角度来看，**FP5** 第二阶段的项目更注重工业应用，尤其是在医学和生物技术方面的应用。

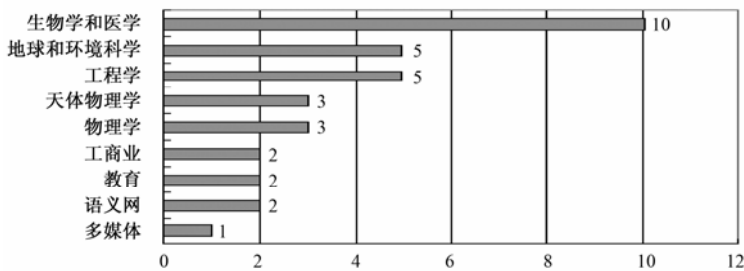


图 4-16 FP5 项目按应用领域分布

4.5 小 结

目前，欧盟第五框架计划下的 **e-Science** 项目已经结束，而第六框架计划下的研究项目也已经全面启动。总的来说，欧盟 **e-Science** 研究主要有以下特点：

（1）欧盟早期的 **e-Science** 研究主要集中在 **e-Science** 底层的基础设施和网格中间件，而最近启动的研究项目则更多地研究 **e-Science** 的高层应用，尤其是在医学和生物技术等特定应用领域的项目较多。

（2）欧盟 **e-Science** 研究的参与机构主要有科研机构、高校院系和企业，而且这三类机构在参与 **e-Science** 研究的数目上基本相当。可见，欧盟 **e-Science** 研究一开始就注重企业界的参与，并始终强调 **e-Science** 的工业化应用。

（3）欧盟 **e-Science** 研究有统一的资助和管理机构，欧盟第五框架和第六框架计划下的网格项目由信息社会与媒体总理事会下的分理事会统一资助和管理。同时，欧盟项目的资助除了来自欧盟委员会外，欧洲的工业界也负担了相当一部分资助额。

（4）欧盟 **e-Science** 研究项目往往是多国合作共同参与，所有欧盟项目至少有来自两个国家的机构共同参与，最多一个项目有 27 个国家参与。此外，欧盟的项目还广泛与欧洲之外的其他国家进行合作。

（5）欧盟内部国家之间的 **e-Science** 研究实力也存在差异，其中按参与项目数目

来统计，排在前五位的国家分别是德国、英国、意大利、法国和西班牙。可见，这些国家是欧盟内 e-Science 研究的主要力量。

（6）在欧盟资助的 e-Science 项目中，资助金额在 100 万到 500 万欧元之间的项目数目最多，资助年限在两到三年的项目最多。从欧盟第六框架计划资助额将超过第五框架计划的两倍可以看出，e-Science 研究在欧盟研究中的重要性在增强。

（7）欧盟早期的网格研究主要集中在分布式超级计算、按需计算和数据密集型应用方面，而欧盟最近启动的网格项目则更加侧重虚拟组织的协调合作应用研究、面向知识发现的应用研究等方面。

（8）欧盟在 12 个网格技术开发领域中开展了网格研究，其中排在前五位的研究领域依次为网格应用、发布、测试平台与网络、数据管理、应用环境与工具、监测与绩效。在这前 5 个领域中，每个领域都至少有 10 个以上项目进行研究。

第5章 亚洲e-Science发展现状分析

5.1 中国大陆e-Science发展情况

中国政府将科技创新作为国家可持续发展的基本保证。改善科研环境,促进科技创新,是中国政府当前的优先考虑。在此背景下,作为一种全新的科研模式,网格乃至 e-Science 对于中国科研的信息化有着特殊重要的意义。目前,中国正在以五大网格建设项目为先导,以三十个相关网格项目为支撑(详见附录 E),以国家科技基础条件平台建设为顶层规划,逐步深入、完善中国的 grid 和 e-Science 基础设施建设。

为了迎接网格的浪潮,我国政府和企业开始重视网格建设。国家高性能计算环境(HPCE)是中国第一个国家级的网格项目,建设时间为 1999 年到 2000 年¹,由国家 863 计划资助(超过 5 千万人民币),由 CERNET(教育网)和 CSTNET(科技网)上的九个不同站点组成,提供对九个不同站点上超级计算机的共享操作,目标是建设面向科学研究的计算网格。但是,HPCE 依然存在一些需要解决,如数据和其他资源的共享能力弱;缺乏协调调度、协作能力弱;真实的网格应用少,只提供了对远程超级计算机的操作;缺少开发环境,网格应用开发困难。

2002 年,科技部公布了 863 网格专项²,投资高达 3 亿元,主要研制面向网格的万亿次级高性能计算机、具有数万亿次聚合计算能力的高性能计算环境;开发具有自主知识产权的网格软件;建设科学研究、经济建设、社会发展和国防建设急需的重要应用网格;制定若干与网格相关的国家标准,参与制定国际标准,使一批发明专利和软件获得受理和登记,形成自主知识产权。

我国在网格领域已有一定的基础和重大的应用需求。为了进一步加强国家的网格建设,国家 863 计划、自然科学基金委、教育部和上海已经或正在部署高性能计算(网格应用)、教育应用和行业应用,同时开展一批亟待解决的共性关键技术的攻关和行业应用示范。目前,我国正在进行的网格项目包括:中国国家网格(CNGrid)、863 空间信息网格、国家自然科学基金委(NSFC Grid)、中国教育科研网格(ChinaGrid)、上海信息网格(ShanghaiGrid)五大网格项目³。此外,由航天二院

¹ 徐志伟,李伟.织女星网格的体系结构研究[J].计算机研究与发展.2002(8):923-929

² 刘鹏.我国网格研究现状[EB/OL].

<http://grid.hust.edu.cn/cgcl/documents/Grid/The%20research%20conditions%20of%20Grid%20in%20China.pdf>. [2007-6-11]

³ 杜渐.网络技术等待商业应用突破[EB/OL].2006-3-10.

<http://www.istis.sh.cn/list/list.asp?id=2657>. [2007-6-11]

和清华大学共同开展的“仿真网格”以及由中科院计算所开发的“织女星网格”也在不断完善之中。

5.1.1 中国国家网格（CNGrid）

1. 概述

2002 年，由国家 863 计划重大专项支持、建设周期四年的中国国家网格（下文简称CNGrid）作为HPCE项目的延续正式启动。该项目总投资 3500 万美元，地方政府、应用承担单位和工业界配套 2 到 3 倍的建设经费，由中科院网络中心、清华大学、上海超级计算中心、北京应用物理与计算数学研究所、国防科技大学、中国科技大学、西安交通大学、香港大学、中科院计算技术研究所、江南计算技术研究所等单位联合承建¹。

CNGrid是聚合了高性能计算和事务处理能力的新一代信息基础设施的测试平台²。通过资源共享、协同工作和服务机制，有效支持科学研究、资源环境、先进制造和信息服务等应用。以技术创新，推动国家信息化建设及相关产业的发展。为强化基础架构建设，中国政府初期投资资金高达 1 亿人民币。

目前，CNGrid已经建立了 11 个应用网格³，即气象网格、资源环境网格、航空制造网格、科学数据网格、新药研发网格、生物信息网格、仿真应用网格、教育网格、城市交通信息服务网格、国家地质调查网格、森林资源与林业生态工程网格等。

国家 863 计划重大专项支持的中国国家网格⁴，是聚合了高性能计算和事务处理能力的新一代信息基础设施的测试平台。通过资源共享、协同工作和服务机制，有效支持科学研究、资源环境、先进制造和信息服务等应用。以技术创新，推动国家信息化建设及相关产业的发展。为此，CNGrid将自身的建设目标定位为⁵：

- ① 建设新一代国家信息基础设施测试平台——中国国家网格，提供高性能计算、资源共享、协同工作的能力；
- ② 在科学研究、环境资源、制造业、服务业中建设若干大型行业应用网格；
- ③ 研制面向网格计算，具有良好的应用开发环境的高性能计算机，装备网格结点，促进我国高性能计算机的研究和产业化；
- ④ 在网格体系结构、网格软件、网格应用技术、网格服务模式、网络安全以及

¹ 中国国家网络首页[EB/OL]. <http://www.cngrid.org/index.htm>. [2007-6-11]

² 陈萍. 北京大学网格计算科研与应用[EB/OL]. 2005-11.

http://conference.jucc.edu.hk/notes/2.17_Professor_Chen_Ping.pdf. [2007-6-11]

³ 中国国家网格（China National Grid）[EB/OL]. http://www.cngrid.org/cn_introduce.htm. [2007-6-11]

⁴ 中国国家网络首页[EB/OL]. <http://www.cngrid.org/index.htm>. [2007-6-11]

⁵ 晓齐. 网格：IT 技术新热点[J]. 信息安全. 2003(3):59-59

网络管理和运行机制等方面突破一批关键技术；

⑤ 在网络技术方面达到世界先进水平，大幅度地提高我国的综合国力和国际竞争能力。

2. 建设重点

- 当前，CNGGrid的建设重点包括¹：
- ① 研制具有每秒 4 万亿次运算能力、面向网格的高性能计算机；
 - ② 建设一个具有 5 万~7 万亿次聚合计算能力的高性能计算环境即中国国家网格；
 - ③ 开发一套具有自主知识产权的网格软件；
 - ④ 建设若干个科学研究、经济建设、社会发展和国防建设急需的重要应用网格；
 - ⑤ 形成若干网格技术的国家标准，参与制定国际标准。

3. CNGGrid体系结构

在建设的过程上，CNGGrid采用三层体系结构的模式（详见图 5-1），强化自身的基础架构建设。其中，资源层主要包括各种网格资源，如主机、程序、仪器、文件系统、数据库等；系统层则包括安全服务、作业服务、用户管理、数据服务、网格管理、网格监控和用户环境，除了安全服务，系统层的其他服务均以信息服务为支撑，以网络软件为实现基础；应用层处于资源层、系统层之上，包括各种计算、数据、软件等在内的各项应用均属于应用层的范围²。

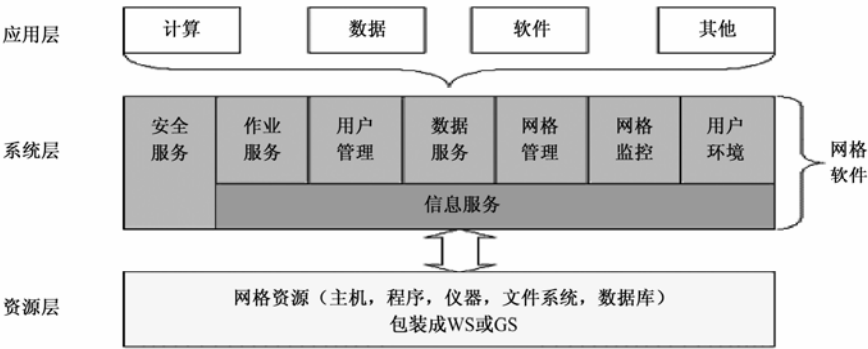


图 5-1 中国国家网格体系架构

¹ 刘路沙.中国：不能错过网格的机遇[EB/OL].http://www.ncic.ac.cn/news/news_08.htm.[2007-6-11]

² 钱德沛.中国国家网格进展[EB/OL].2004-4-29.
<http://www.chinagrid.net/grid/paperppt/bigfile/Conference/IDG2004Spring/K1.ppt>. [2007-6-11]

4. 管理运行机制分析

CNGrid 由以下单位共同承担建设（详见表 5-1）。

表 5-1 CNGrid 承担机构一览表

机 构 名 称	机 构 性 质	参 与 人 数（人）
中科院网络中心	科研机构	7
上海超级计算中心	科研机构	2
北京应用物理与计算数学研究所	科研机构	1
江南计算技术研究所	科研机构	1
中科院计算技术研究所	科研机构	1
国防科技大学	大学	1
中国科技大学	大学	1
清华大学	大学	3
香港大学	大学	2
西安交通大学	大学	1

为保证日常研究与建设的顺利开展，CNGrid项目组成立了由科技部直接负责的领导小组（CNGrid管理架构详见图 5-2），对国家网格的运行进行政策性规范和指导¹。领导小组任命国家网格专家组，负责国家网格技术的研发规划，指导网格的部署和运行管理。中国国家网格设立网格运行支持中心（简称运行中心），负责国家网格的技术支持与维护，保证国家网格的正常运行，推动网格在中国的应用与发展。运行中心接受专家组的指导。中国国家网格设有核心结点和普通结点，称为国家网格的成员。核心结点保证长期为国家网格提供资源与服务。网格成员应当遵守国家网格管理办法，完成各项工作。

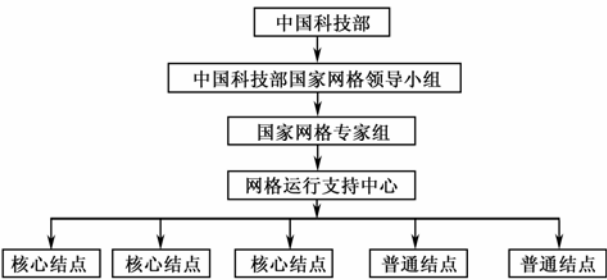


图 5-2 CNGrid 管理架构

¹ 中国科技传播网[EB/OL].<http://book.netpub.com.cn/rssfeed.asp?cid=2>. [2007-6-11]

5. 国际合作

中国国家网格与英国grid计划有着十分相近的目标和研究内容,双方已经开展了有效的密切合作。2005年11月15日,中英两国政府签署了在grid和网格研究方面开展合作的部长级合作备忘录¹,为两国进一步拓展合作渠道,扩大合作规模,开展实质性合作研究奠定了基础。在2005年10月14日于法国举行的GRID@WORK大会上,中国国家网格与其他国际网格项目密切合作,成功为大会举办的第二届Plugtests国际网格大赛提供了网格设施服务,在网格计算能力的支持与服务方面,排名第二²,成为迄今中国最大规模网格设施参与实质性国际合作的领头羊,这次国际合作提升了国家网格的组织协调与国际合作能力。

5.1.2 上海信息网格 (ShanghaiGrid)

1. 概述

上海网格(下文简称ShanghaiGrid)是上海市科委重大科技攻关项目,是中国“五大网格”之一,于2003年启动。ShanghaiGrid整合上海市已有的高性能计算能力、海量存储能力,连接了上海主要的网格计算研究机构(包括上海交通大学、同济大学、上海大学、上海超级计算中心、复旦大学),形成了聚合计算能力6000亿次/秒、聚合存储能力4TB的上海网格测试平台,凝聚了30位教授、150名博士后、博士生及硕士生的研究队伍,上海市科委总共投资900万人民币,并带动国家及地方资金5000万³。

2. 建设目标

ShanghaiGrid的主要目标是在两年内,依托国家863计划、国家自然科学基金、国家教育部和地方重大科研攻关项目的最新研究成果,研究符合国际标准的信息网格体系结构和关键技术,开发具有自主知识产权和推广价值的信息网格系统软件、应用开发环境和虚拟研究平台,通过对虚拟组织中的计算、数据、软件等各种信息的共享和协同,建立具有上海特色的信息网格,并初步实现交通信息网格典型应用⁴。

¹ 刘斌.中国国家网格正式开通[EB/OL].2005-12-23.

<http://www.biotech.org.cn/news/news/show.php?id=29484>. [EB/OL]

² 中国国家网格运行管理支持中心.中国国家网格参加法国 Grid PlugTests 大赛[EB/OL].

2005-10-21. http://www.cngrid.org/02_news/news/cn_news_news_051021.htm. [2007-6-11]

³ 上海交通大学.谭宗仁博士受聘上海网格总体组顾问[EB/OL].

2004-6-25. <http://www.sjtu.edu.cn/newsnet/newsdisplay.php?id=1029>. [2007-6-11]

⁴ Advances in ShanghaiGrid (in Chinese)[EB/OL].

<http://unpan1.un.org/intradoc/groups/public/documents/APCITY/UNPAN016937.pdf>. [2007-6-11]

3. 建设重点

ShanghaiGrid第一阶段的建设重点包括以下四个方面¹：

① 研究信息网格基础设施的环境、协议和标准。这包括：信息网格关键互联参考协议；网络结点管理软件、网络结点接入软件；网络问题求解工具、网络结点性能模型与评价；网络结点接入软件的实现；网络结点的性能评价；网络结点的安全与可靠性；

② 开发信息网格的系统软件，建立4个主要的内部网络结点。这主要包括两方面的内容：网络结点建设，即上海交大、同济大学、上海大学、上海超级计算中心四个网络结点的建设；网格系统软件开发，主要涉及数据传输、访问与集成、资源管理与发现、网格服务中间件、网络安全服务和信息网格开发环境等；

③ 开发基于P2P的虚拟研究平台。研究P2P数据管理和信息检索技术；研究P2P环境下的即时消息通信；研究P2P环境下的资源共享；研究P2P环境下的信息发布；实现一个基于P2P技术的虚拟协作平台；

④ 研究基于网络的交通拥塞控制与疏导应用。这主要包括：遵循OGSA的交通信息服务需求分析；动态交通网格模型与优化试验；区域混合交通流并行仿真研究；基于海量交通信息的路况预测；交通多源数据融合与集成技术；交通出行方案服务点播与信息展示。

4. 体系结构

ShanghaiGrid的体系结构实质上是一个通用的五层网络结构，它包括构造层、连接层、资源层、聚合层和应用层。如图5-3所示，ShanghaiGrid构造层主要由各个结点的资源组成，目前已经建立了五个结点；连接层主要包括各类连接协议，涉及互联结构的规划与设计，为应用层的各类应用得以实现提供互联基础；资源层负责管理各个结点的资源，并保证数据的正确传输；聚合层由网格服务中间件、资源发现与网格监控和数据共享访问与集成三部分组成；在聚合层之上为应用层，目前包括了ShanghaiGrid提供的三种具体应用，即基于网络的交通信息发布、基于网络的交通拥塞控制与疏导、基于网络的交通状况模拟。为实现这三种应用，需要创建一个信息网格开发环境，并利用安全服务技术、P2P虚拟研究平台技术、信息技术领域的网格服务、出行方案的建模与优化技术、混合交通流的并行仿真技术以及交通路况预测方法提供相应的支撑。简单地说，ShanghaiGrid的体系结构由三部分组成，底层是信息网格结构与各类协议，中间层是信息网格的各种关键技术，而顶层却是信息网格的各种应用。随着建设的深入，ShanghaiGrid的应用领域将进一步得到拓展。

¹上海交通大学.谭宗仁博士受聘上海网络总体组顾问[EB/OL].

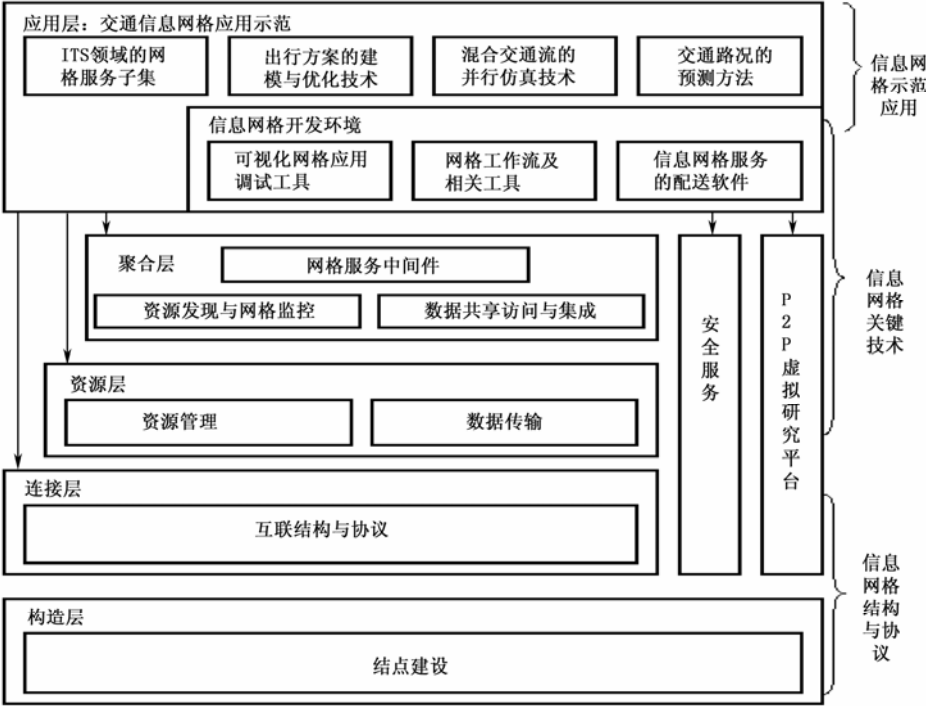


图 5-3 上海网格的四层体系结构

5. 管理运行机制分析

ShanghaiGrid 第一阶段项目总投资 2800 万人民币,共由 10 个承担机构(表 5-2)。

表 5-2 上海网格承担机构一览表

机 构 名 称	机 构 性 质
上海交通大学	大学
同济大学	大学
复旦大学	大学
上海大学	大学
上海超级计算中心	科研机构
上海交通信息中心	政府机构
华东计算机技术研究所	科研机构
IBM	公司企业
Intel	公司企业
Platform	公司企业

目前, ShanghaiGrid 的四个内部网格节点分别是上海交通大学、同济大学、上海大学 and 上海超级计算中心。为保证建设的如期进行, ShanghaiGrid 项目组将 ShanghaiGrid 分为总体组和分项目组, 整个项目由上海市科委直接领导, 设立首席专家一名, 总体组设有专家组, 共由来自项目承担机构中的 7 位专家组成。此外, 项目设总体组顾问一名。在总体组下面, 上海网格又分为四个子项目组对项目实施管理(图 5-4)。

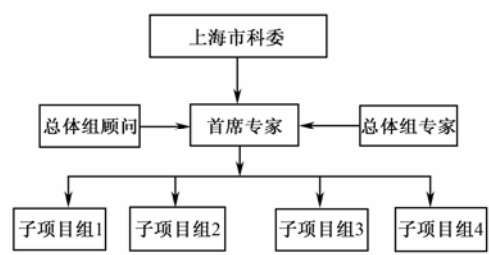


图 5-4 上海网格管理架构

6. 国际合作

自投入建设以来,Shanghaigrid 就非常注意国际合作与交流。2003 年,Shanghaigrid 在国际网络会议 GCC 展示研究成果, 得到国际学术界的一致好评。

5.1.3 中国教育科研网格 (ChinaGrid)

1. 概述

中国教育科研网格ChinaGrid计划是教育部“十五”211 工程公共服务体系建设的重大专项¹, 于 2003 年启动。中国教育科研网格侧重于计算网格。教育部的规划希望通过中国科研网格的建设可以为科研提供服务, 该项目的建设侧重点就是计算网格。但同时, 为了服务于教育, 中国教育科研网格也预期在第二期部署远程教育网格, 这也是第二期的重点之一。

2. 建设目标

中国教育科研网格将充分利用中国国家教育科研网CERNET和高校的大量计算资源和信息资源², 开放相应的网络软件, 配合网络计算机的使用, 将分布在教育和

¹ 中国教育科研网格.ChinaGrid 简介[EB/OL].
<http://www.chinagrid.edu.cn/chinagrid/index.jsp>. [2007-6-11]

² hp 中国.ChinaGrid 计划全面启动[EB/OL].
http://www.hp.com.cn/enterprise/success/server/051116_02.asp. [2007-6-11]

科研网上自治的分布异构的海量资源集成起来，实现CERNET环境下资源的有效共享，消除信息孤岛，提供有效的服务，形成高水平低成本的服务平台，将高性能计算送到教育和科研网用户的桌面上，成为国家科研教学服务的大平台。

3. 建设重点

中国教育科研网络的建设重点分为两个阶段。第一阶段侧重于服务科研，第二阶段的建设重点是部署远程教育网格。在第一阶段中，中国教育科研网络的建设重点体现在以下四个方面¹：

① 在 12 所具有超级计算环境的高等院校建立 ChinaGrid 网格主结点，每个主结点的计算能力至少达到每秒 5000 亿次、存储空间至少达到 5TB，最终形成聚合计算能力超过每秒 6 万亿次量级、总存储空间超过 60TB 的教育科研网格；

② 开发一整套 ChinaGrid 公共支撑平台软件（CGSP），作为整个 ChinaGrid 的网格中间件。争取在网格计算的基础研究和应用研究方面走在世界前列；

③ 提供面向五类教育及科学研究应用的专业网格及其开发环境，这五类专业应用网格包括图像处理网格、生物信息学网格、大学课程在线网格、海量信息处理网格和计算流体力学网格；

④ 加强网格学科建设，为高等院校培养和储备一批高素质的网格研究及网格应用人才。

中国教育科研网络的长远目标是建设中国乃至世界上“最大、最先进、最实用”的网格系统。从最初的 12 所高校开始，中国教育科研网络的覆盖面正在不断扩展之中，最终将覆盖全国的大、中、小学和其他科研教育机构，其用户群也会由现在的领域专家和专业用户扩展为包括中小学生在内的普通用户。

4. 体系结构

中国教育科研网络的体系结构分为三层，自下而上分别为高性能计算环境（校园网格）、中国教育科研网络公共支撑平台（CGSP）和具体应用层。

① ChinaGrid 公共支撑平台（CGSP）

中国教育科研网络公共支撑平台（CGSP）是中国教育科研网络为各种网格应用提供基础支撑的网格核心中间件。CGSP 的第一期开发工作已经完成，第一版本已于 2005 年 1 月正式发布，这是全球第一个基于目前 WSRF 技术开发的网格支撑平台。CGSP 是国际上第一个遵循 OGSA 架构，参照 WSRF 规范实现的网格中间件。它提供了一套完整的网格服务支撑平台，对教育和科研系统中的各种资源进行整合，屏蔽网格资源的异构性和动态性，为各种科学计算与工程研究提供高性能的、高可靠

¹ 金海. ChinaGrid 建设目标[EB/OL].2005-5-30.<http://www.cutech.edu.cn/jiaoyuxinxihua/000132.asp>.
[2007-6-11]

的、安全方便的透明网格服务，形成一套公共网格服务体系。目前已经有生物信息学、图像处理、计算力学等的应用在该平台上调试通过。

CGSP共划分为六个功能模块¹：

网格门户（Portal）：作为中国教育科研网络的网格服务展现方式，网格门户是最终用户使用网格的入口。

网格开发环境：其主要功能是提供资源网格化封装的工具包和网格构建管理工具包，以及提供面向网格环境的编程模型，用于复杂网格作业的开发。

信息中心：其主要功能是负责网格环境中各类资源信息的管理，实现一个全局的资源视图，提供网格信息服务，并实时更新网格资源信息。

统一管理层：其主要功能是为网格环境中各类作业操作提供基础支持。

网络安全：所提供的功能包括用户身份的认证、资源和服务的授权、加密传输以及用户身份到资源授权的映射等。

计算结点层：真正提供网格服务的结点，可以是集群等大型计算设备，也可以是普通微机和工作站等。

CGSP是一组互相配合的软件组件，包含若干个可独立运行的软件，用于支持开发过程、运行过程、系统安装过程和系统管理的各个环节。

② 具体应用层

中国教育科研网络的应用层包括五大具体应用，分别是图像处理网格、生物信息学网格、大学课程在线网格、海量信息处理网格和计算流体力学网格。

5. 管理运行机制分析

参与中国教育网格项目第一期建设的高校包括华中科技大学²、清华大学、北京大学、北京航空航天大学、上海交通大学、华南理工大学、东南大学、西安交通大学、东北大学、中山大学等12所高校，目前已经发展到了20个主要大专院校。

中国教育科研网格计划自筹备起就受到教育部领导的高度重视，专门成立了以教育部赵沁平副部长为组长的领导小组和一个由12位国内从事网格研究及其网格应用的专家组成的专家组。专家组设组长一名，副组长两名。专家组下面又按应用类型分为五个项目组，分别是图象处理网格项目组、生物信息学网格项目组、大学课程在线网格项目组、海量信息处理网格项目组和计算流体力学网格项目组（图5-5）³。

¹ 中国教育网络. ChinaGrid 建设目标：最大、最先进、最实用[EB/OL]. [2008-10-15]

http://www.edu.cn/zhong_dian_ke_ti_5168/20060517/t20060517_178950.shtml

² ChinaGrid 公共平台建设[EB/OL].<http://grid.hust.edu.cn/platform/html/chinagrid.html>. [2007-6-11]

³ ChinaGrid 公共平台建设[EB/OL].<http://grid.hust.edu.cn/platform/html/chinagrid.html>. [2007-6-11]

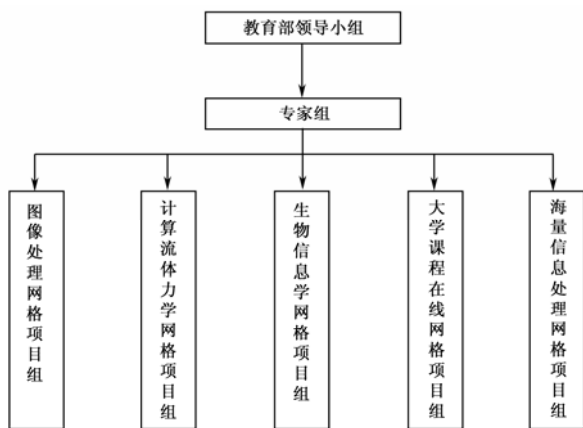


图 5-5 ChinaGrid 管理架构

6. 国际合作

在国际合作方面，中国教育科研网络主要通过人员交流与项目合作等方式与外界建立联系。如它与美国、新加坡共同参与 AccessGrid 项目的建设；与欧盟开展合作；同时与 Globus 联盟签署三年期的合作协议；除此之外，它还与产业界开展密切合作。如与惠普公司在网络监测方面进行合作，与 IBM、Sun 等业界巨头联手合作。

5.1.4 国家自然科学基金委网络建设项目（CROWN）

1. 概述

现代科学研究活动对计算技术提出新要求：科研资源总量大，但缺乏共享和有效的利用；科研合作需求广，但缺乏便捷有效的协作环境；建立以网络为基础的科学活动环境成为国际计算机技术研究和应用的热点和前沿领域。在此背景下，国家自然科学基金委于 2004 年开始启动 NSFCGrid 建设计划，该计划预计建设时间三年，目标是建立一个基于网络的科研环境。目前在建的项目是 CROWN（China Research and Development Environment over Wider-area Network，CROWN），从 2004 年起投入建设，持续时间三年¹。

CROWN 由北京大学、北京航空航天大学、中国科学院网络中心、清华大学、国防大学共同参与建设。截至 2004 年 12 月²，CROWN 的一期试验网络连接了包括北京航空航天大学、中国人民解放军国防科学技术大学、中国科学院计算机网络信息中心、北京大学、清华大学等国内重要科研单位，跨越北京、长沙的异地试验网络，

¹ 基于广域网络的科研环境[EB/OL].<http://www.crown.org.cn/>. [2007-6-11]

² 基于广域网络的科研环境[EB/OL].<http://www.crown.org.cn/>. [2007-6-11]

并通过中国科学院计算机网络信息中心同环太平洋网络中间件与应用联盟（PRAGMA）、中美俄先进应用环球网络（GLORIAD）等全球范围的广域试验网络相连接。

2. 建设目标

作为与国家自然科学基金委网格计划相配套的建设项目，CROWN将自身目标定位¹，建设一个以网络为基础的科学活动环境综合试验平台，为国内各个领域的科学家提供一个综合的网络科研环境。

为保证目标的顺利实施，CROWN 将自身的建设目标进一步细分为两个方面：一方面，通过建设使自身成为各领域科学家开展各自领域科研活动的工具和平台，成为各类科研资源的统一访问门户；另一方面，作为国家自然科学基金委研究专项的试验环境，它期望作为承载相关科研活动的基本测试环境，验证相关理论和技术的正确性和效率。

在此基础上，CROWN制定了自身的目标架构图（如图 5-6 所示）²。透过该图，可以清晰地看到，在上述目标的驱动下，建立具有国际水准的以网络为基础的科学活动综合试验平台、完成以科研资源共享和科研协同为重点的若干个示范应用、取得具有我国自主知识产权的理论成果和专利技术、发表一批高水平的学术论文和学术专著、形成一支高水平的研究队伍、提高我国在网络计算环境领域的整体创新能力和国际竞争力是CROWN的建设宿命³。

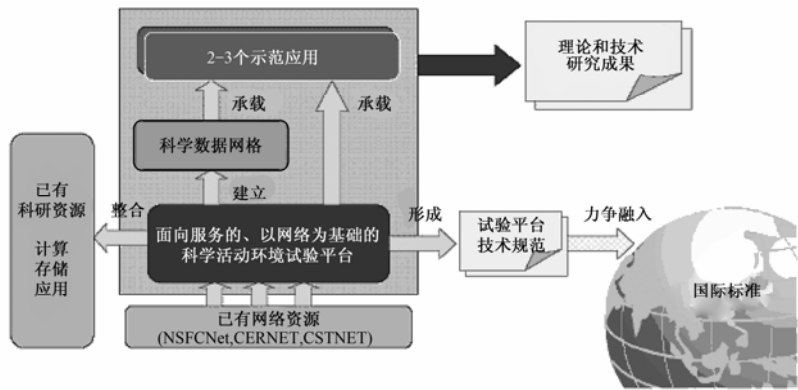


图 5-6 CROWN 目标架构图

¹ 关于 CROWN[EB/OL]. http://www.crown.org.cn/aboutus/about_us.jsp. [2007-6-11]
² 怀进鹏. 网络计算环境中间件技术研究 2005 年度检查报告-总体进展[EB/OL]. 2005-12-19.
<http://www.crown.org.cn/download/download.jsp?iDownloadType=2>. [2007-6-11]
³ 怀进鹏. 网络计算环境中间件技术研究 2005 年度检查报告-总体进展[EB/OL]. 2005-12-19.
<http://www.crown.org.cn/download/download.jsp?iDownloadType=2>. [2007-6-11]

3. 建设重点

在建设的过程中，CROWN将它期望构建的网络科研环境细分为三大层次¹。第一，这种网络科研环境需要在现有网络基础设施之上构造一个由广域网络及若干骨干的资源结点构成的全国范围的试验网络；第二，它需要针对不同科研领域对网络科学活动的共性需求，提供一组基于服务的网格计算中间件平台及一组辅助工具，以简化不同领域、不同用途的网格应用的开发和管理；第三，它需要建立一组典型示范应用以验证平台的可用性，建立网格应用的开发和管理规范。

在此基础上，CROWN 制定出适合营造这种科研环境的基本设计思路，采用面向服务的体系结构，将科研资源服务化，并将服务化的资源按照其物理隶属关系组织起来，在已有的 TCP/IP 网络互联的基础上构造可控可管的资源层叠网（Resource Overlay），实现资源（服务）能力的描述与发布、自动的服务发现和服务交互，在广域网络中实现对资源能力的共享和综合利用。

截止于 2005 年，CROWN 已经在香港科技大学、国防科技大学、北京大学、清华大学、中国科学院网络中心、北京航空航天大学、中国科学院大气物理研究所大气科学和地球流体力学数值模拟国家重点实验室、重庆大学建立了八个分布式测试平台。2006 年，CROWN 的建设重点是在西安交通大学、中国科技大学、上海超级计算中心、上海交通大学四个单位布置测试平台，进一步扩展 CROWN 的建设规模，提升其服务能力。

4. 体系结构

从图 5-7 可以看出，CROWN 采用三层结构的模式进行建设。自下而上，这三层分别是资源服务层、应用支撑层、用户支撑层。图 5-7 表明，资源服务层由计算服务代理、数据服务代理、存储服务代理、应用服务代理等代理组成，为了与计算资源、数据资源、存储资源以及各种应用资源建立代理联系，需要由相应的适配器协助工作，这一层的各类服务代理由服务容器进行聚合。

CROWN 的应用支撑层由共性服务与领域服务两大块组成。其中，共性服务作为领域服务的基础设施，需要承担保证领域服务正常开展的各项任务，如任务调度、流程管理、数据管理、注册与发现、元数据管理、服务合成等。依靠共性服务的保障，CROWN 便可以提供面向科学数据网格、大气监控与分析、高能物理、生物计算等各种各样的领域服务，并通过门户管理服务、可视化服务将各种领域服务提供给

¹ 怀进鹏. 网络计算环境中间件技术研究 2005 年度检查报告-总体进展[EB/OL].2005-12-19.

科学工作者使用¹。

为了保证用户支撑层、应用支撑层和资源服务层各层功能的顺利实现与正常发挥，需要由启用开发者采用各种核心技术，如新型体系结构、服务封装与管理、共享与协同、安全与管理、面向领域的应用开发技术等，辅以必要的开发方法与工具进行设计与创建；另外，为了保证系统的正常运行，需要由平台管理者在安全方面进行维护、监督与管理。

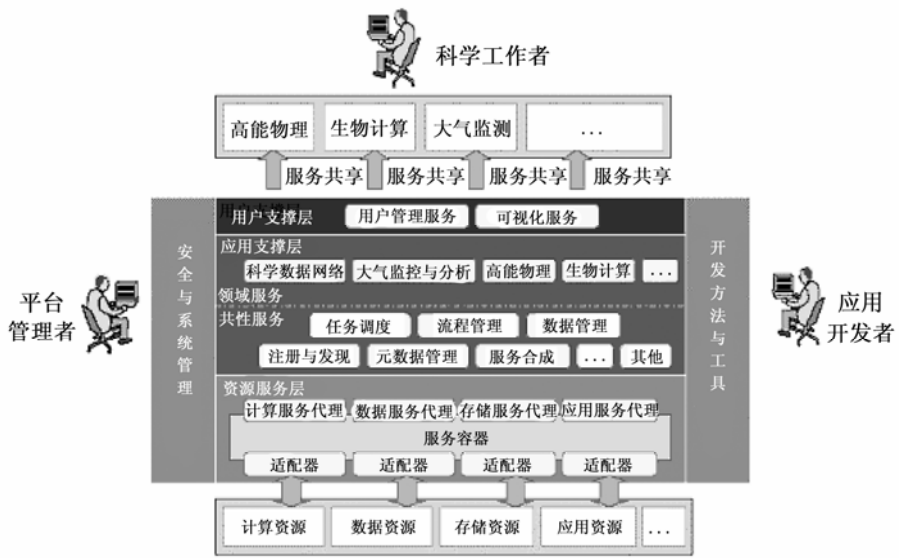


图 5-7 CROWN三层体系架构²

5. 管理运行机制分析

由北京航空航天大学牵头承建，各参加单位依托自身的建设优势，承担相应的建设模块，如北京航空航天大学在中间件建设方面富有经验，就负责 CROWN 中间件的建设，由中国科学院网络中心负责科学数据网络数据访问服务系统软件的设计开发，由国防科技大学承担服务组合支撑平台的构建，等等。

在建设的过程中，由北京航空航天大学协调、部署五个参建机构的各项任务分配，并对国家自然科学基金委负责，如图 5-8 所示。

¹ 怀进鹏.网络计算环境综合试验平台—北航进展汇报[EB/OL].2005-12-19.
<http://www.crown.org.cn/download/download.jsp?iDownloadType=2>. [2007-6-11]

² 怀进鹏.网络计算环境综合试验平台—北航进展汇报[EB/OL].2005-12-19.
<http://www.crown.org.cn/download/download.jsp?iDownloadType=2>. [2007-6-11]

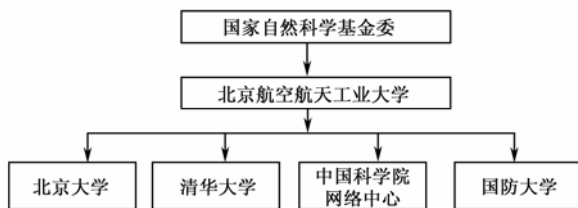


图 5-8 CROWN 管理架构

6. 国际合作

① 与国外重点研究机构建立合作关系。与英国 grid Engineering Task Force(ETF)合作。开展 CROWN 系统评估，作为中国第一个被评估的网格中间件，GT4 于 2005 年 6 月从系统管理（服务部署）、网络安全等角度接受 EFT 的评估。此次评估中，共有三所 grid 研究中心参与。即南安普敦地区 grid 中心（设立在南安普顿大学）、伦敦地区 grid 中心（设立在帝国理工大学）以及东北地区 grid 中心（设立在纽卡斯尔大学）。此外，CROWN 项目组与 OMII（Open Middlware Infrastructure Institute, OMII，开放中间件基础架构研究所）建立互访制度。

② 与国外重点大学建立合作关系。与英国里兹大学合作（Leeds University）合作，共同创建 CoLaB（Cooperation of Leeds University and BeiJing Univarsity）。召开英国七次专题研讨会，四次在英国进行，三次在中国进行。

③ 积极参与 AHM2005 会议。项目组有关人员在会上作了 gViz 可视化应用的报告。

④ 合作建设项目。参与 OMII-Europe 项目建设，与 MOU 签署协议，进行网格中间件和可信服务相关方向的合作研究与技术开发；与 OMII 一起开发网格中间件；北航与欧洲及国内多家合作伙伴共同承担欧盟 FP6EU-China 网格项目，将开展中国的网格环境与欧盟 EGEE 互联互通的工作；与 Globus 一起参与 OMII-Euro 项目；与微软软件亚洲研究院共同参与可信计算的项目建设。

⑤ 系统代码开源：与国际著名开放源代码组织 OjectWeb 联合。在世界著名的开放源代码网站上发布有关系统的源代码，提高研究团队参与国际技术合作的力度。

⑥ 建立走出去请进来的合作机制。项目组参与 9 次国际会议特邀报告、组织或参与国际会议近 10 次；参加 INFORCOM、ICPP、VLDB、APPT、grid 2005AHM、iGrid2005、ICEBE2005、MoDel、ICRE、ICRI、SCC2005 等国际会议并宣读论文；四名成员到英国 OMII 总部进行学术交流；与法国国家计算机与自动化学院 INRIA 建立了 7 个研究组。在人员交流方面。原英国 e-Science 首席科学家、e-Science 核心计划主任 Tony Hey、OMII 副所长 Steven Newhouse 教授、Globus 负责人 Ian Foster 教授、W3C 主席及办公室主任等 grid 相关领域的杰出专家到项目组开展交流。在机构合作方面。与美国的 Globus 联盟、英国的 grid 建设机构(OMII,ETF)、法国的 INRIA、德国的 JSI 以及中国香港科技大学开展合作。

5.1.5 863 空间信息网络

1. 概述

863 空间信息网络项目 (SIG) 于 2001 年启动, 建设周期四年, 由中国地质调查局、中国地质调查局发展研究中心、国防科技大学、国家信息中心、武汉中地公司、环境监测院、地质科学院水文地质研究所、航天航空大学、网络公司 9 家单位共同承建¹。

空间信息网络 (SpatialInformationGrid, SIG) 是一种汇集和共享地理上分布的海量空间信息资源, 并对其进行一体化组织与协同处理², 具有按需服务能力的空间信息基础设施。它关注的重点是知识获取与处理, 利用汇集的各种空间信息资源、统一的空间信息管理和处理平台、应用以及相关工具, 在空间信息网络中, 各种空间信息资源被统一管理和使用, 空间信息处理是分布式协作的和智能化的, 用户可以通过空间信息网络门户透明地使用整个网络上的各种资源。空间信息网络的最终目标是把 Internet 上的空间信息服务站点连接起来, 实现服务点播 (Service On Demand) 和一步到位的服务 (One Click Is Enough)。发展空间信息网络将从空间信息应用与服务的技术体系和基础高度角度推动我国空间信息资源的共享与应用, 满足日益增长的多层次、多样化空间信息应用需求, 这对提升我国空间信息基础技术和应用的水平, 促进国民经济和社会发展, 具有十分重要的意义。

2. 建设目标

“十五”国家 863 计划信息获取与处理技术主题的战略目标是面向我国信息资源设施建设的重大需求³, 发展高分辨率多维空间信息获取技术, 研制高分辨率机载光学、微波对地观测数据获取系统, 开发轻小型星载高空间分辨率多光谱成像仪, 掌握自主的先进小卫星对地观测系统技术。系统性地开展对地观测数据定量化、智能化处理技术研究, 突破大型地理信息系统、空间信息网络化共享中的关键技术, 通过重大应用示范, 构建我国的空间信息网络, 推动我国空间信息资源的产业化。

¹ 刘恕. 让地质空间信息自由流动[EB/OL].

http://www.863.org.cn/863_105/news/news_infotech/200507150001.html. [2007-6-12]

² 空间信息网络概述[EB/OL].

<http://www.nudt.edu.cn/newweb/intercommunion/whatissig.htm>. [2007-6-12]

³ 863 计划信息获取与处理技术主题组织部分优秀成果参加第 18 届国际卫星对地观测委员会 (CEOS) 全会展览会

[EB/OL]. 2004-11. http://www.most.gov.cn/kjbgz/200411/t20041126_17530.htm. [2007-6-12]

空间信息网格具体应用研究空间信息网格在信息地质中的应用¹，例如，空间信息网格在地质环境仿真、地震预报、油藏模拟、矿产勘探等领域中的应用；空间信息网格在数字城市中的应用，例如，数字城市空间信息服务集成，特别是城市突发事件应急响应，如火灾、地震、洪水、台风等；空间信息网格在电子政务中的应用，如空间信息网格在辅助政府决策方面的应用，包括城市规划、重大工程选址等。空间信息网格在数字流域中的应用，如利用空间信息网格进行整个流域的数字模拟，包括洪水演进、流域生态、坝位选址等。

3. 建设重点

863 空间信息网格项目的建设重点包括以下三方面²：

① 空间信息网格框架体系和关键支撑技术主要研究空间信息网格的体系结构及其应用环境基础框架，实现海量空间信息在线分析处理和服务。基于空间信息标准和协议，建立空间信息网格核心技术验证平台，提供空间信息处理和大型地理信息系统的基本测试环境。结合资源环境等领域空间信息与服务体系的建设，提供关键技术支持。研究内容包括空间信息网格体系结构；空间信息网格的资源描述、组织与管理技术；空间信息网格中的在线分析处理和服务技术；空间信息处理和大型地理信息系统测试环境；空间信息网格核心技术验证平台和示范应用。

② 基于空间信息网格框架的城市空间信息应用服务系统在城市空间信息基础设施的基础上，选择重点示范城市，根据城市建设、管理与产业发展需求，基于空间信息网格框架，建立城市空间信息资源汇集与共享的基础平台，开发面向政府决策和社会化服务、具有基本按需服务能力的运行系统，促进城市空间信息应用与产业化发展。研究内容包括城市空间信息共享应用的政策与协议；基于空间信息网格框架的城市空间信息获取、处理与应用服务系统的总体设计，并建立相应的软硬件平台；城市多源空间信息获取、共享、传输等关键技术以及基于空间信息网格框架的城市空间信息应用服务系统开发。

③ 关键技术研究。空间信息网格关键技术研究空间信息网格关键技术包括超大规模空间数据的存储与管理、多源空间信息资源共享与融合技术、空间数据元数据及其服务技术、空间信息网格服务技术（包括网格资源信息服务、网格性能信息服务、网格服务信息服务等）、空间信息系统之间的协同工作问题、空间信息网格环境下的地理空间数据可视化等。

¹ 金江军.网格技术在地球信息科学中的应用[EB/OL].2004-2-25.

http://industry.cidnet.com/art/322/20040225/92292_1.html. [2007-6-12]

² SIG 空间信息网格[EB/OL].http://grid.hust.edu.cn/article_63.html. [2007-6-12]

4. 体系结构

空间信息网络体系结构主要由资源层、服务层和应用层组成（详见图 5-9）¹。

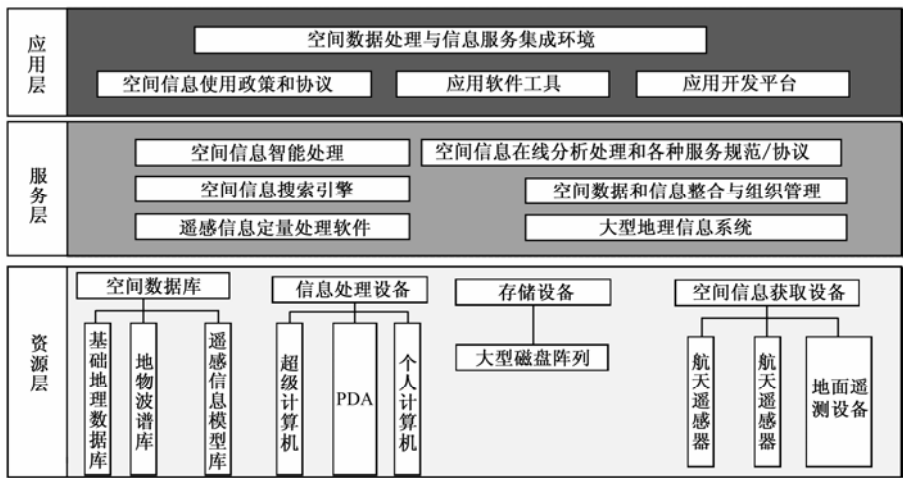


图 5-9 863 空间信息网络体系结构

① 资源层：这些资源包括已建和在建的各类空间数据库（包括基础地理数据库、地物波谱库和遥感信息模型库）、各种信息处理设备（包括超级计算机、PC、PDA）、各种存储设备（大型磁盘阵列）、各种空间信息获取仪器（包括航空/航天遥感器、地面遥测设备），它们通过 Internet 或各种无线通信设施实现物理连接。

② 服务层：提供一个空间信息一体化管理与处理平台，为综合使用各类资源提供数据存储、组织管理、分发、检索和处理等服务，主要包括遥感信息定量处理软件、大型地理信息系统、空间信息搜索引擎、空间数据和信息整合与组织管理、空间信息智能处理、空间信息在线分析处理和各种服务规范/协议等。

③ 应用层：提供一个面向应用领域的空间信息集成应用环境，即在上述空间信息网络服务层的基础上，面向各个具体应用领域，对空间信息的使用模式和使用特点，提供空间信息使用政策和协议、应用软件工具、应用开发平台等，建立空间数据处理与信息服务集成环境。

5. 管理运行机制分析

目前，863 空间信息网络由中国地质调查局、中国地质调查局发展研究中心、国防科技大学、国家信息中心、武汉中地公司、环境监测院、地质科学院水文地质研

¹ SIG 空间信息网络[EB/OL].http://grid.hust.edu.cn/article_63.html. [2007-6-12]

研究所、航天航空大学、网络公司 9 家单位组成，由 863 计划办公室统一进行管理（见图 5-10）。

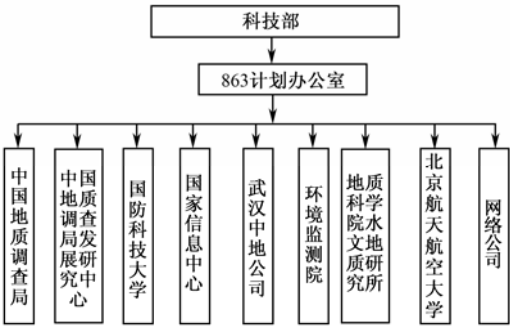


图 5-10 863 空间信息网格管理架构

6. 国际合作

在项目研究阶段，研究单位 4 次组约 10 人次到英国、加拿大、澳大利亚、德国、丹麦就数据模型、地质科学数据共享、标准等方面进行了交流与讨论。

5.2 中国台湾e-Science发展情况

在中国台湾科学会支持下，中国台湾大学物理系、“中央大学”物理系与“中研院”物理所团队于 2001 年起开始参与 LHC 试验。并于 2002 年获得“中央研究院”网络运算团队支持，开始共同导入 LCG（LHC Grid Computing）。在数年的合作努力下，如今获得 CERN 与 WLCG 国际合作团队的认可，受邀担任目前亚洲的 Tier-1 中心。

在 WLCG 的基础与经验下，“中央研究院”网络运算团队正在积极建立新一代的研究架构，以支持更多大规模的基础研究与跨领域的应用（如数位典藏、大气科学、生物资讯等），并提供参与国际合作的有效管道。此外，藉由 WLCG 的国际合作，“中央研究院”网络运算团队于 2004 年 12 月正式加入欧盟支持的 EGEE（Enabling Grids for eScienceE）计划。该计划整合欧盟各国研究团队及计算资源，建立一个全球性的网格运用环境。藉由 WLCG 基础架构的导入，EGEE 计划业已成为引领全球性网格中介软件的发展，以及推动 e-Science 与产业应用的重要驱动力。成为 EGEE 的成员，“中央研究院”网络运算团队能有效掌握并直接参与 e-Science 基础架构之发展，提升对相关研究之支持能力，并维持与国际间一流团队之长期紧密合作。

King Project（Knowledge Innovation National Grid），即知识创新网格，是中国台湾发展 e-Science 的重要措施之一，它主要以网络为核心平台，通过网络及相关技术，连接各种分布在各地的资源，使之成为一个统一的平台系统。该项目由中国台湾高

速网络计算中心承担，计划历时 4 年，投资 3000 万美元完成。

2005 年，中国台湾“中央研究院”与欧洲粒子物理研究中心(CERN)签署 WLCG (World-Wide LHC Computing Grid) 备忘录。该备忘录的签署使中国台湾正式成为全球 LHC 网格计划中亚洲 Tier-1 中心，并与全球其他十个 Tier-1 中心（即加拿大 TRIUMF、法国 CC-IN2P3、德国 GridKA、意大利 CNAF、荷兰 NIKHEF/SARA、北欧 NDGF、西班牙 PIC、英国 RAL、美国 BNL 及 FNAL）共同整合资源与服务，并建立沟通协调机制。此外，中国台湾“中央研究院”与其他五个中心（CC-IN2P3, CERN, CNAF, RDIG, RAL）共同担任 WLCG 全球核心维运中心与咨询服务中心。

5.3 韩国e-Science历史与现状

韩国从 2001 年开始，在国家层面上积极与美国、英国、日本等发达国家以及中国等发展中国家开始合作，共同推进 e-Science 的建设。

K*Grid 是韩国的当家网格项目。它主要由 MIC (Ministry of Information and Communication)管理，由 KISTI(Korea Institute of Science and Technology Information)资助，总费用大概 3200 万美元，为期 5 年（2002-2006 年）。K*GRID 是韩国国家网格架构的一个重要组成部分，主要由韩国网格论坛、K*GRID 基础架构建设、网格中间件技术和网格应用研究 4 个部分构成。K*GRID 的主要目的是为工业和科学研究建立一个强大的、拥有大量计算能力的研究环境。

据韩国科技部透露，截至 2010 年，韩国政府将投入 1027 亿韩元，建设新一代信息化科研环境——e-Science。e-Science 项目于今年底正式启动，2010 年完成。

5.4 日本GRID历史与环境

1. ITBL

ITBL (IT Based Laboratory) 项目在 2001 年 4 月通过立项。它的目标就是通过信息技术来构建一个虚拟的共同研究环境。日本政府将在一定期限内为该项目投入 1 亿 6 千万欧元。在不久的将来，ITBL 将通过 SuperSINET 将日本国内 100 或者更多的超级计算机有效地连接起来，从而使得被连接的这些超级计算机可以共享资源，同时通过这样，来建立一个依靠大规模实验设备的因特网从而得以被人们共同使用的系统，等等。

此外，ITBL 将开发出并向人们提供用于使用数据库和超级计算机所需要的软件，促进研究类型的多样化，并提高包括超级计算机在内的各种设备的使用效率。目前，ITBL 有六个成员研究所。他们分别是国家材料科学研究所 (NIMS, National Institute

for Material Science)、国家地球科学与灾难预防研究所 (National Research Institute for Earth Science and Disaster Prevention, NIED)、日本国家航空宇宙实验室 (National Aerospace Laboratory of Japan, NAL)、物理化学研究所 (RIKEN, Institute of Physical and Chemical Research)、日本原子能研究所 (Japan Atomic Energy Research Institute, JAERI)、日本科学与技术振兴事业团 (Japan Science and Technology Corporation, JST)。在这些研究所当中, 原子能研究中心和物理化学研究中心负责处理实施 ITBL 计划所必须共用基础技术的开发。此外, 各个研究所还将独立承担实施 ITBL 整个实施过程所需的软件的开发。

根据 e-Japan 集中计划, 筹建 ITBL 的目标就是到 2005 年实现对日本国内所有研究所的超级计算机的共享, 为了达到这一宏伟目标, 该项目制定了以下措施:

- 依靠日本国内的各研究所的计算机资源资源网络, 开发出能够实现安全、简单的共享的技术, 并对之加以示范;
- 通过高速网络, 筹备和共享日本各研究所的计算机资源;
- 使得广大范围内的研究者能够获取到技术和资源, 从而实现各个前沿科技领域中研发技术的创新。

2. Grid Data Farm

日本的 Data Farm 网格项目, 主要用于 Petabyte 数据量的高能物理实验数据的分析和处理, 与欧洲数据网格相连。东京工业大学的松岗聪教授 (全球网格论坛指导委员会成员) 认为, 到 2004 年, 网格技术将成为日本信息技术领域的基础设施类大项目。

3. NAREGI (National Research Grid Infrastructure)

该项目于 2003 年启动, 由 NII 和 Institute of Molecular Science 发起, 由国立情报学研究所日本的文部科学省 (Ministry of Education, Culture, Sports, Science and Technology) 资助, 共投资 1.25 亿美元, 其主要目的是为了建立一个能够符合国际标准的, 在高级科研和教育领域能够提供广域分布的强大计算环境。

4. Business Grid

该项目由日本经济产业省资助, 为期三年 (2003—2005 年), 一共投资 26 亿日元。该项目的主要目标是建立下一代的商业应用框架, 得到了很多 IT 巨头企业的支持, 如日立, 富士, NEC 等。该课题主要解决三方面的问题, 一是资源可视化; 二是自动基于策略的管理; 三是商业应用配置。

5.5 小 结

亚洲网格建设起步较晚，各国几乎是在进入 21 世纪之后才逐步建立起自己的国家网格。由于是后发项目，亚洲网格计划的目标都比较明确，以应用为主，其目的都是尽快提高科研或者工业效率。

亚洲的网格项目都是作为国家重点项目而存在，由一定的政府机构牵头，采取项目申请的方式，来组织科研机构和大专院校进行研究建设。其经费一般也由政府部门出资，但是由于日韩都拥有着国际 IT 巨头，他们对利用网格快速提高工作效率和加快研究项目开发效率很感兴趣，所以他们在网格开发计划中无论是研发还是投资都占有很重要的地位。相比之下，中国的网格计划中，民间研发机构是凤毛麟角，投资更是不必多说。从硬件设施上来看，亚洲网格环境并不亚于欧美等地区和国家。亚洲网格都没有单纯依靠 GT 系统来进行网格建设，而是基于 GT 开发网格中间件等系统。

5.5.1 各个项目承担主体分析

1. 中国

中国的五大网格项目中，由 863 发起的项目有 2 个，另外三个项目的发起者分别是教育部，NSFC 和上海市政府，由此可见中国的 e-Science 项目主要是由政府主导和资助，同时依靠企业自愿加入，保证资金。通过对可以查到的项目预算的分析（863 空间信息网格和 ChinaGrid 的预算信息暂缺），中国对五大网格项目的总投入大约为 3 亿 7900 万人民币，其中政府直接投资达到了 12900 万人民币，约占总投资的 34.1%，其他部分都是由政府投资所带动的其他投资，大部分来自于当地政府、应用组织和相关企业。

由于中国国情决定，中国的五大网格项目的发起者都是政府机关，CNGRID 共有承担单位 8 个，其中研究机构 3 个，教育机构（大专院校）5 个，有两个主结点分别建立在北京中国科学院网络信息中心和上海超级计算中心；NSFC 具体承担情况暂缺；shanghaiGRID 的具体项目承担者中，研究机构两个，教育机构 4 个，企业 3 个；ChinaGrid 由 20 个院校共同承担；空间信息网格的具体承担情况暂缺。由已知的数据分析，中国三大网格项目（CNGRID,SHANGHAI GRID,CHINA GRID）承担机构共有 28 个，科研机构 3 个，占 10.7%；企业 3 个，占 10.7%；教育机构 22 个，占 78.6%。由此可见，在中国网格建设的主体是各大大专院校，而且北京航空航天大学，清华大学，西安交通大学，长沙国防科技大学，中国科技大学，上海交大，同济大学和复旦大学都同时担当了两个或以上国家五大网格的具体项目承担者。

2. 日韩

和中国类似，日本和韩国的主要 Grid 计划也是由政府主导，比如日本的文部科

学省和经济产业省,以及韩国的 KISTI。但是资金的承担情况和中国却有很大的差别。由于韩国与日本都有着 IT 巨头的存在,因此他们的 grid 建设目标更现实。在日本的两个 grid 项目中, Business Grid 就是面向企业的应用 Grid,其研究主体也是公司中的研究机构;而另外一个 NAREGI,则是由 4 个科研单位和 4 所高校联合开发进行的。韩国的 K*GRID 的项目中,拥有超级计算机和结点计算机的单位共有 14 个,包括 8 所大学和 6 个科研单位。

5.5.2 亚洲e-Science的技术框架与技术模块分析

1. 技术模块

通过对亚洲几个网格项目的结构分析,发现亚洲的 grid 项目技术部分都是一种类似的三层结构。本节将按照这种三层结构来分析亚洲网格的技术模块。

第一层是物理层。物理就是 Grid 架构中存在的物理设备,包括高性能计算机、数据库、存储设备和高性能网络,以及所共享的软件及文件等。现在亚洲拥有自己的网格项目的国家,比如中国、日本、韩国、新加坡,都已经拥有了自己的高性能计算机和大量的网格信息资源。比如中国的联想深腾 6800,每秒 5.3 万亿次浮点运算,拥有 80TB 磁盘阵列存储容量。

第二层是网格中间件。中日韩三国都有自己开发的网格中间件, CNGRID 有自己独特的网络中间件,即中国科学院软件所开发的 VEGA GOS; CHinaGrid 有自己的 CGSP;韩国的 K*GRID 开发的中间件结构 MoreDream,日本的 NAREGI Middle Ware。他们虽然组成不尽相同,但是都有类似的结构:

- (1) 网格连接模块。该模块的主要目的是与物理资源进行连接。涉及的问题是解决网格的结构与协议;
- (2) 资源连接模块。主要涉及数据管理和数据传输控制。很多中间件都利用 globus toolkits 中的 Data Management 模块相应的 api 进行开发;
- (3) 服务与任务管理模块。主要是处理用户的各种服务请求和为用户的请求分配任务,并且进行管理和处理;
- (4) 网格应用模块。在此基础上要进行应用网格开发;
- (5) 安全模块。处理用户的认证、映射等;
- (6) 信息中心模块。这是网格中间件的核心部分,主要任务是协调网格中间件其他各部分的良好运行。

第三层是网格应用层。就是利用网格中间件所提供的应用网格接口应用网格开发的成果。

当前, Globus Toolkits 是全球最通用的网格操作系统,这个由全球网格论坛 (GGF) 下属 Globus 项目组成员联合开发的 Globus Toolkit 标准工具包,已被公认为当前建立网格系统和开发网格软件事实上的参考标准。Globus 项目是国际上与网格计算相关的最有影响力的项目。但是值得一提的是,中国的两大网格都自行开发了能够与 GT 相匹敌的 Gos,分别是 CNGRID 的 VEGA 织女星 GOS 和 ChinaGrid 的 CGSP

(ChinaGrid Supporting Platform)。

2. 织女星

织女星网格体系结构的基本思想是把网格看成一台虚拟的超级计算机系统。与现有的计算机系统类似，它包括硬件、系统软件和应用三个组成部分。

织女星网格操作系统主要分为 4 个功能模块，分别是 GOS 应用、GOS 系统服务、GOS 核心服务、GOS 主环境。

3. CGSP

ChinaGrid 开发了一个为各种网格应用提供基础支撑的网格核心中间件——ChinaGrid 公共支撑平台 (CGSP)，并于 2005 年 1 月发布了 v1.0。很多专家对 CGSP 非常看好，认为完全与 Globus Toolkit 4 有得一比，并受到了 Globus 项目的重视。如前文所述，CGSP 共划分为六个功能模块。CGSP 是一组互相配合的软件组件，支持像 ChinaGrid 网格的应用的开发、调试、部署、运行管理以及系统监控等各个环节。基于 CGSP，整个 ChinaGrid 网格系统可以构造成一个分层树型结构（见图 5-11）。

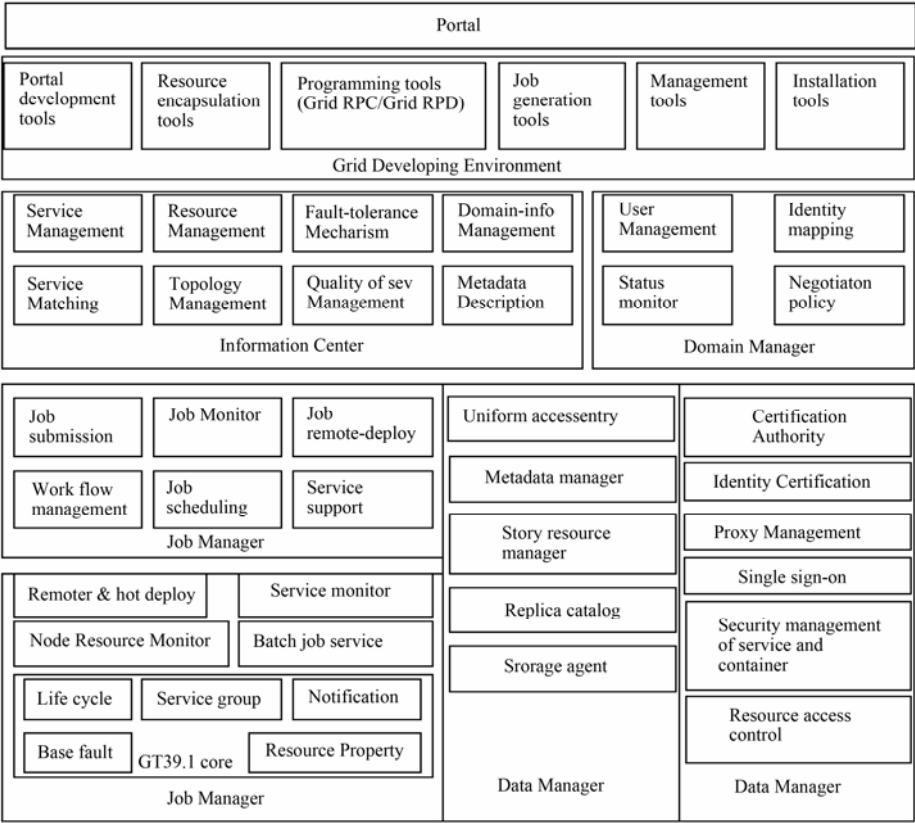


图 5-11 CGSP 框架

第二篇 关键技术篇

英国e-Science Core计划的项目带头人Tony Hey认为，e-Science或eResearch的核心是构建e-Infrastructure，而e-Infrastructure的目标是开发和利用先进的信息技术来创建一个信息基础架构，以支持多领域的和协作的科学研究与创新，它需要解决：对研究数据的生产、采集、掌管、保存和分析问题；数字环境下被研究对象的建模和仿真问题；动态的、分布式的、虚拟的研究团队的组建和工作模式问题¹。

为解决以上问题，需要借助网格。网格作为运行于高性能网络之上的一系列中间件，通过它可以组合来源于多个领域的专家和资源，促进资源的共享，提供一个动态的多机构参与的虚拟组织以解决上述问题。

综上所述，e-Science 的关键技术主要包括以下 4 个方面：

- (1) 基于网格的资源和服务共享技术；
- (2) 研究数据的采集、管理、保存和分析技术；
- (3) 研究对象的建模和仿真技术；
- (4) 动态的、分布式的、虚拟的研究团队的组建和协同技术。

本篇将基于以上分析，从网格技术、数据管理、对象仿真建模和虚拟研究团队构建四个方面，分析当前 e-Science 的若干关键技术问题及其解决方案，力图反映当前 e-Science 的技术进展。

¹ Towards an e-Infrastructure for Research and Innovation: A Progress Report on e-Science [EB/OL].
(2008-8-20), <http://www.nesc.ac.uk/events/ahm2004/presentations/TonyHey.ppt>

第 6 章 基于网络的资源和服务共享技术

网络是 e-Science 的基础框架。通过网络技术,可以将高速互联网、高性能计算机、大型数据库、传感器、远程设备等融为一体,将分散在不同地理位置的电脑组织成一台“虚拟的超级计算机”,实现计算资源、存储资源、数据资源、信息资源、软件资源、通信资源、知识资源、专家资源等的全面共享。

本节在对网络技术发展历程进行总结的基础上,对网络的主要技术和标准进行解析,分析与网络密切相关的信息技术,并对一些基于网络的工具和系统进行介绍。

6.1 网 格 概 述

6.1.1 网络的概念和主要类型

网络是借鉴电力网的概念提出来的,网络的最终目的是希望用户在使用网络计算能力时,如同现在使用电力一样方便。人们在使用电力时,不需要知道它是从哪个地点的发电站输送出来的,也不需要知道该电力是通过什么样的发电机产生的,不管是水力发电,还是通过核反应发电,人们使用的是一种统一形式的“电能”。同样,网络也希望给最终用户提供的是与地理位置无关、与具体的计算设施无关的通用的计算能力。

应当指出,网络仅在使用方便这一点上类似于电力网络。由于信息资源的存储与服务系统、资源的描述与表达、提供服务的方式千差万别,信息服务和计算网格与电力网格技术在复杂度上有很大的差别。网格系统及其应用的目的在于对位置分散、异构和动态变化的虚拟机构的资源和服务进行集成管理,一旦需要就可以对这些计算机、应用服务、数据和其他资源进行访问。实现这一目标的关键是实现标准化,将地理位置分散、属于不同机构的资源和服务视为单个虚拟系统,进行有机接入。

迄今为止,网格的发展经历了三个阶段。

第一阶段是网格的萌芽阶段。这一阶段开始于 20 世纪 90 年代初,研究内容是关于千兆网试验床以及一些元计算方面的工作。

第二阶段是从 20 世纪 90 年代中期到末期的试验阶段。这一阶段出现了一些比较重要的开创性和奠基性的研究项目,比如 I-WAY, Globus, Legion 等。

第三阶段的发展始于 21 世纪初。21 世纪以来是网格计算迅速发展的阶段,关于网格的研究、开发和应用的项目大量出现,出现了影响很大的组织——全球网络论

坛（Global Grid Forum, GGF），同时网格计算也不再仅仅局限于科学界，工业界与学术界联盟，更多的网格研究爱好者正致力于使网格计算在更广泛的领域得到推广和应用。

但是到目前为止，对“什么是网格”这一问题还存在不同的认识。被称为网格之父的IanFoster在《什么是网格？判断是否网格的三个标准》一文中，提出有必要明确网格的定义¹。他认为网格已经从晦涩的科学研究走向了广泛的应用。目前已知的网格有：计算网格、数据网格、科学网格、访问网格、知识网格、生物网格、传感器网格、集群网格、校园网格、Tera网格以及社区网格。如果在局域网上配置了任务调度程序，就建成了一个“集群网格”。这同时引出了一系列的问题。比如在同一个网络上的网络文件系统（Network File System, NFS）是否就算作“存储网格”；甚至，具备了处理器、内存、硬盘和网卡的工作站，是不是也算是一个“PC网格”；什么计算机系统不能算作网格。

IanFoster认为网格现在的情形与Internet在20世纪90年代早期的情形相似，那时供应商声称诸如SAN和DECNET之类的私有网格也是Internet的一部分，另一些人则认为每个局域网都是一种形式的Internet。只有当广域网、局域网都广泛采用TCP/IP协议之后，这种概念的混乱才得以澄清。

关于网格的原始概念在20世纪60年代就开始出现。Len Kleinrock在1969年就曾提出，将来可以看到“计算机工具”的广泛应用，就像现在的电力和电话工具一样，这种“计算机工具”将会在整个社会中为居家和办公提供服务。

1998年，IanFoster和Carl Kesselman在《网格：21世纪信息技术基础设施的蓝图》（The Grid: Blueprint for a New Computing Infrastructure）一书中给出了计算网格的定义：计算网格是一系列的软、硬件基础设施，这些基础设施借助某些可靠的、一致的、普遍的和经济的模式，能够实现高性能计算的性能。网格是构筑在互联网上的一组新兴技术，它将高速互联网、高性能计算机、大型数据库、传感器、远程设备等融为一体，为科技人员和普通老百姓提供更多的资源、功能和更强的交互性。互联网主要为人们提供电子邮件、网页浏览等通信功能，而网格功能则更多更强，人们可以透明地对其他资源进行使用、计算、存储等操作。

2000年，IanFoster和Steve Tuecke在其合著的《网络剖析》（The Anatomy of the Grid）一书中，对原来定义进行了精炼，提出网格计算涉及“在多机构组成的动态虚拟组织中提供协调一致的资源共享和问题解决”。其中关键的概念是在一组参与者（包括提供者和消费者）中协调资源共享的能力，并为某种应用目标使用结果资源池。他们提出：“人们所关心的共享不仅仅只是文件交换，而更应该是对计算机、软件、数据和其他资源的直接访问，这是解决工业、科学和工程领域出现的一系列的

¹ Ian Foster, What is the Grid? A Three Point Checklist. [EB/OL]. (2008-8-20),
<http://www-fp.mcs.anl.gov/~foster/Articles/WhatIsTheGrid.pdf>

协调问题和资源代理策略所必需的”。这种共享必须是高度可控制的，资源提供者和消费者必须明确定义什么能共享、允许谁共享、在什么情况下共享。共同遵循这个共享规则的个人和组织的集合被称为“虚拟组织”。

对于什么是网格，福斯特提出了三个简单的衡量基准：

(1) 能够协调非集中控制的资源

网格集成和协调存在于不同控制领域中的资源和用户。如用户的桌面与中央控制计算机、同一公司的不同管理单元或不同的公司。网格可以处理出现在这种环境中的安全、政策、支付和成员资格等问题。如果不具备这些功能，那它实际就相当于一个本地管理系统。

(2) 使用标准、开放、多用途的协议和界面

网格是建立在具有认证、授权、资源发现和获取的多用途协议和界面之上的。这些协议和界面的标准化和开放很重要的，否则，它只是一个特殊的应用系统，通用性不高。

(3) 提供高质量的服务

网格允许以协调一致的方式使用资源，以提供各种高质量的服务，如响应时间、吞吐量、有效性、安全和为满足复杂用户需求的多种类型资源的协调分配。

这三条基准在科学界所进行的各种大型的网格项目中都是适用的，如国际上的“Data Grid”项目，包括 GridPhyN (Grid Physics Network), PPDG (Particle Physics Data Grid), EU DataGrid、iVDGL (international Virtual Data Grid Laboratory)、DataTAG 等的分布式数据处理系统，美国国家航空航天局 Information Power Grid 将荷兰的五个大学链接起来的 Distributed ASCI Supercomputer (DAS-2) 系统，美国能源部的 DOE Science Grid 和 ASCI DISCOM Grid，以及正在进行的将美国主要研究机构站点链接起来的 TeraGrid。每个系统都是将不同机构的资源整合起来，使用开放、多用途的协议 (Globus Toolkit) 来协调和管理共享，多方位地提供高质量的服务 (安全、可靠和性能)。

网格的主要优点在于：高可用性、高可靠性、高系统性能、便于模块化及增量扩展、自动实现负载和资源共享、对暂时的高负荷能够得到及时响应、容易实现功能扩展。

这些优点在科学研究中则表现为：提高了超级计算的可获得性，实现了计算的平民化，一般人也能够实现超级计算；通过非凡的计算能力，可以发展一些随机性的推理科学，“资源都在那里，让它跑跑看，我们能够得到什么”；能够实现功能组装，进行灵活的计算。

6.1.2 网格在科学研究中的应用类型

有的人认为网格的界定并不取决于其体系结构，而是它所能提供的应用、商业

价值和科学结果。

近年来，网格技术备受关注，高能物理、生物信息、地球观测、全球变化以及数字保存等领域都在积极探究网格技术的应用，以期更有效地实现数据的搜寻、整合、分析及保存、管理，实质而有效地增进国际间分工合作与资源整合，共同处理基本粒子物理、生物学的基因体与蛋白质体乃至宇宙起源等更大尺度的科学问题；突破空间障碍，建立彼此可共享信息、资源、应用工具和知识的虚拟组织，通过虚拟组织则进一步可以灵活地形成各种不同的合作机制；有效管理分散于各地、各机构的异质资源，促进资源共享与利用；整合网络上的各种计算资源和数据资源，协助科学家处理更大尺度、更高复杂度的问题；建构安全稳定的数据存储与共享机制。

网格应用能够在研究手段、研究规模、研究深度等方面促进以下几种类型的重要科研迅速进步和提高。

(1) 计算密集型应用：交互式模拟(如气候模型)，大规模的仿真和分析(如星系的构成，战场仿真)，工程计算(如参数计算)。

(2) 数据密集型应用：实验数据分析(如高能物理)，图像和传感器分析(天文、气候、生态研究)。

(3) 分布式协作型应用：联机仪器使用(如显微镜、X 射线设备等)，远程可视分析(气候研究、生物学)，工程技术领域(如大规模结构测试，化学工程)。

上面所举的例子，都是当前“大科学”的一些例子，非常需要实现不同组织内科学家的协作，实现计算资源、数据和仪器设备的共享和协同。

与上述分类方法不同，研究人员根据计算角度和应用特性，将网格分为：计算网格、存储网格、数据网格、信息网格、知识网格和各类应用网格等。

6.1.3 网格技术的研究重点

无论用什么方式去描述网格系统，都必须清醒地意识到：网格是计算机网络环境中实现用户访问地理位置分散、异构的计算机系统资源的一种通用应用服务平台。因此，它本身研究的重点是如何在计算机网络提供的已有服务平台之上，实现方便的协同工作与资源共享，而不是研究计算机网络技术本身。

综观当前的网格研究，其重点集中于信息网格的体系结构、网格标准、资源表示、统一认证、安全技术等。

1. 网格体系结构

随着网格计算研究的深入，人们越来越发现网格体系结构的重要。网格体系结构的研究涉及如何建造网格的技术，包括对网格基本组成部分和各部分功能的定义和描述，网格各部分相互关系与集成方法的规定。显然，网格体系结构是网格的骨架和灵魂，是网格最核心的技术，只有建立合理的网格体系结构，才能够设计和建

造好网格，才能够使网格有效地发挥作用。

2. 网格的标准协议

网格协议不但需要开放、多用途，而且需要标准化。正是标准化的协议使网格实现了任何参与方之间动态的资源共享，而标准化协议作为实现多用途服务和工具提供的手段也是很重要的。

给出标准的“InterGrid”协议的定义是当前网格所面临的重要问题。在协议标准化方面，全球网格论坛发挥着重要的作用；而在应用实践方面，有网格计算事实上的标准 Globus Toolkit。在全球网格论坛的推动下，定义了 OGSA (Open Grid Services Architecture, 开放式网格服务体系结构)。OGSA 在 Globus Toolkit 协议基础上进行了改进和扩充，以满足新的需求。除此之外，还有 OGSi (Open Grid Services Infrastructure, 开放式网格基本结构)、WSRF (WebServices Resource Framework, Web 服务资源框架) 等协议也在制定和完善之中。

3. 网格中间件

网格计算的核心构件是网格中间件，它通过聚合计算设备、高性能存储器、数据库，甚至包括科学仪器，来为用户提供对计算力 (Computing Power) 随时随地的、透明的、远程的、安全的、可靠的访问。就像电网把电力提供给墙上的每个插座，网格聚合可用的计算资源，然后把计算力提供给网格环境中的每个用户，使人们可以轻而易举地为一些科研工作创建和使用大规模、多学科、动态的、分布式的、高性能的应用环境，如高能物理数据分析、基因信息处理、气候建模、宇宙观测、实时遥感数据分析和数据同步、大型数据集交互分析和虚拟现实可视化等。

4. 资源监测和服务质量

网格是一种分布式计算机环境，它需要多种网络设备、计算机系统、数据资源和支撑服务协同工作，为虚拟组织提供一个可靠的研究环境。为了防止这些设备和服务工作失效，需要对这些设备和服务进行监测。在网格监测方面，标准化和能够对多种异构资源进行监测是其研究重点。

5. 统一认证问题

为了方便用户一次登录就可以利用网格资源，系统需要实现统一认证。

从全球范围来看，随着信息技术和网格技术的发展，为了支持各种网络服务，统一用户认证的需求被各国相关领域广泛关注。从目前的发展来看，可以把统一认证管理分为二个层次：第一层是机构间的统一认证管理系统，这种模式可以很方便地扩展到国家范围内使用，比较著名的系统是来自美国 Internet2 的 Shibboleth，西班牙学院网络提供商 RedIris 的 PAPI 和英国 JISC 的 Athens；第二层是主要应用于单独

机构的认证管理，比较著名的包括耶鲁大学的 CAS，华盛顿大学的 Pubcookie，斯坦福大学的 WebAuth，密歇根大学的 KX.509（X.509 certificates via Kerberos）以及荷兰的 A-Select。

其中，英国已经建立二个全国范围的统一认证管理系统：Athens 实现对 e-learning 资源的管理，e-Science 数字证书方案实现对研究资源的访问。目前 JISC（Joint Information System Committee）在 Core Middleware initiative 项目中开始资助基于 Shibboleth 的国家统一认证计划。

6. 安全问题

信息网格的目标是将 Internet 上提供信息服务的站点连接在一起，让所有用户都可以享受这些信息服务。同时，这些站点又可能分属于不同的组织机构，各组织机构可以独立地管理属于自己的网络结点。在实际应用中，信息网格必须为站点管理者提供访问控制等安全管理机制，管理者可以自由地决定共享哪些信息、与谁共享。当然，这种权限控制必须是易维护且独立于数据源本身的，同时应该在逻辑上提供不同的安全管理层次和控制力度。

6.1.4 网格的基本组件和功能

网格环境的构建层次从下至上依次为：

1. 网格结点

分布在 Internet 上的各类资源组成，包括各类主机、工作站甚至 PC，它们是异构的，可运行在 Unix、Windows 等各种操作系统下，也可以是上述机型的机群系统、大型存储设备、数据库或其他设备。

2. 中间件

网格计算的核心。负责提供远程进程管理、资源分配、存储访问、登录和认证、安全性和服务质量等。

3. 开发环境和工具层

提供用户二次开发环境和工具，以便更好地利用网格资源。

4. 应用层

提供系统能接受的语言，如 HPC++和 MPI 等，可配置其他一些支持工程应用、数据库访问的软件，还可提供 Web 服务接口，使用户可以使用 Web 方式提交作业并取得计算结果。

网络计算环境要求不影响各结点本地的管理和自主性。不改变原有的操作系统、网络协议和服务，保证用户和远程结点的安全性，尽量使用已存在的标准的技术，并能提供可靠的容错机制。为此，网络的基本组件包括：

- (1) 管理体系：管理域按区域层次划分，决定管理信息流的流向；
- (2) 通信服务：随应用目的不同提供不同的服务。包括可靠的点对点和不可靠的组播通信，支持各种通信协议，提供通信链路延迟、带宽和可靠性等指标；
- (3) 信息服务：提供方便可靠的机制，获得不断变化的各结点信息和状态；（见图 6-1）
- (4) 命名服务：提供全局统一的命名服务，典型的有国际通用的 X.50 标准或 Internet 上 DNS 标准；
- (5) 文件系统：提供一个分布式文件系统机制、全局存储和缓存空间；
- (6) 安全认证：这部分应包括登录认证、可信赖、完整性和记账等方面的安全性，这是网络计算的难点。也是系统成败的关键；
- (7) 监视系统：提供监视系统资源和运行情况的工具；
- (8) 资源管理和调度：提供透明的资源调度，高效地利用可利用的资源是系统的核心；
- (9) 资源交易机制：为鼓励不同组织或资源拥有者加入系统，应提供一种计算资源的交易机制，允许提供资源者获得利益，使系统能动态地取得最好的性价比资源；
- (10) 编程工具：必须提供丰富的用户接口和编程环境，提供最常用的语言，如 C、C++。

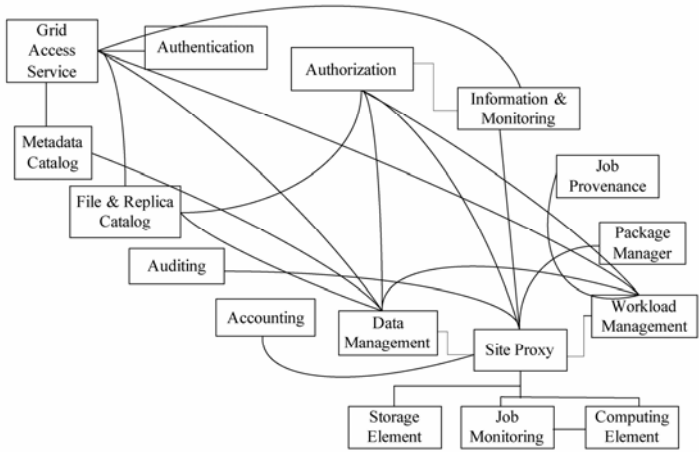


图 6-1 典型的Grid服务图¹

¹ Roberto Barbera, Introduction to the GENIUS portal and the GILDA testbed, [EB/OL]. (2008-8-20), <http://www.dma.unina.it/~murli/GridSummerSchool2004/session-06.htm>

6.2 网格的主要技术标准

国际上主要网格标准化组织包括全球网格论坛 GGF (Global Grid Forum)、OMG (Object Management Group)、W3C、GridForum 以及 Globus.org 等。虽然这些团体都试图争夺网格标准的制定权,但目前大多数网格项目都是基于 Globus Toolkit 所提供的协议及服务,例如,美国的物理网格 GriPhyN、欧洲的数据网格 DataGrid、荷兰的集群计算机网格 DAS-2、美国能源部的科学网格和 DISCOM 网格、美国的 TemGrid 等。Globus Toolkit 已经成为事实上的网格标准。

2000 年底成立的全球网格论坛是一个致力于网格技术标准化的社团。GGF 自成立以来已经先后制订了若干网格工作文档 (Grid Work Document, GWD) 和 45 个“网格最终文档”(Grid Final Document, GFD),其中包括开放式网格服务体系结构 (Open Grid Services Architecture, Version 1.0, OGSA) 和开放式网格基本结构 (Open Grid Service Infrastructure, OGSi)。

2003 年上半年,符合 OGSA 规范的 Globus Toolkit 3.0 正式发布,这标志着 OGSA 已经从一种理念、一种体系结构,进入了实践阶段。作为一种开放架构和开放标准基础设施, Globus Tooltik 提供了构建网格应用所需的很多基本服务,如安全、资源发现、资源管理、数据访问等。

为了使网格技术能够实现远程资源调用和复杂数据的交换, OGSA 和 OGSi 借用了 W3C 组织制订的 Web 服务的相关标准,如 Web 服务描述语言 (WebServices Description Language, WSDL) 和简单对象访问控制协议 (SOAP),并积极参与和推动 Web 服务资源框架 (WebServices Resource Frame, WSRF) 标准的制订,准备作为 OGSi 的替代层。图 6-2 为网格技术标准发展演化的示意图。

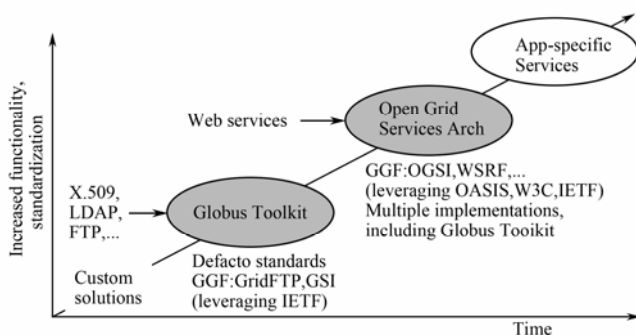


图 6-2 网格技术标准发展演化的示意图¹

¹ Lee Liming, The Globus Toolkit Ecosystem, [EB/OL], (2008-8-20),

<http://www.nesc.ac.uk/events/ahm2004/presentations/T3a.pdf>

6.2.1 开放网格标准体系（OGSA）

1. 概述

OGSA 旨在推动对分布式异构资源的无缝使用和管理。这里，术语“分布式”、“异构”、“资源”都是广义的概念。例如，“分布式”既包括地理上的分布——资源通过某种连接网彼此紧密连接，也包括全球范围内多个组织间资源松散的、间断的连接。“资源”指系统内或系统上操作所需要的任何人工制品、实体或知识。这种基础设施提供的效用被作为一组功能实现，图 6-3 显示了对其中一些功能的逻辑的、抽象的、半层次化的描述。

图 6-3 中，底层是基础资源（base resources），即被一些底层物理或逻辑的实体（元素）支持的资源，其中的物理实体包括 CPU、内存、磁盘等物理元件，逻辑元素指的是许可、内容、操作系统进程等软件。虚拟化是与所要虚拟的实体紧密耦合的，因此用于命名底层实体或元素的术语也被用来命名它们的虚拟对应物。这些基础资源通常是本地拥有和本体管理的，对其的配置和定制也在本地进行，但可以被远程共享。因为实际的实体或元素可能会迅速变化，也可能有多个来源，因此它们所支持的基础资源的特征、服务质量、版本和可用性等方面也会随之变化。

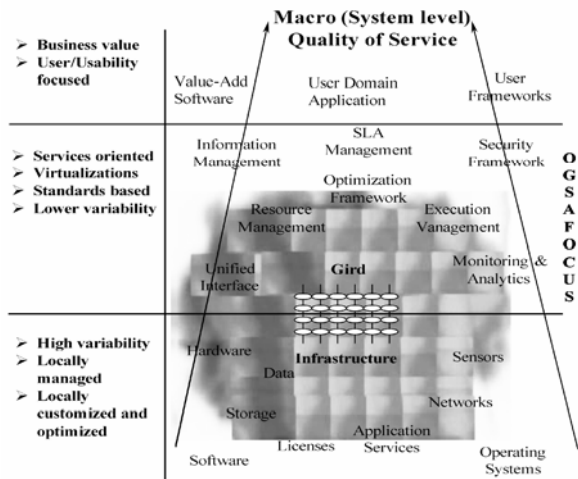


图 6-3 网络基础设施的概念视图¹

需要说明的是，尽管这里单独讨论了基础资源，以便将 OGSA 概念与传统的资源概念相联系，但本篇的后续讨论中并未对基础资源与其他用作资源的服务进行具体区分，而是使用了资源的通用概念。

在基础资源之上的中间层描述了虚拟化和逻辑抽象的一个更高的层次。虚拟化

¹ Global Grid Forum, The Open Grid Services Architecture, Version 1.0, [EB/OL].(2008-8-20), http://forge.gridforum.org/sf/go/projects.ogsa-wg/docman.root.published_documents.ogsa_1_0

和抽象的目的是定义与 OGSA 网络相关的各种能力，这些能力可以单独或组合应用，作为基础设施支持更高层的应用或用户域中的过程。这组能力是相对稳定和标准的，实现或实施这些能力的方式以及它们被用户域中的应用组合和扩展的方式决定了基础设施的宏观质量。需要注意的是，图 6-3 中所描述的功能仅仅是 OGSA 全部功能的一个例子，本篇的其余部分将对更完整的功能集合进行详细讨论。

OGSA 面向服务的本质属性意味着发挥服务作用的虚拟化资源与其他服务在结构中处于同等地位（如中间层或上层中的服务），而这种同等的关系意味着结构中服务之间的交互可以由任何服务启动。此外，中间层中的服务需要使用和管理底层的虚拟资源，利用底层资源传递个别服务将要提供的能力。因此，底层中的实体被看作是与 OGSA 讨论相关的，或者是 OGSA 讨论的一部分。在上层中，对使用 OGSA 能力实现面向用户或领域的功能或过程（如商业过程）的应用及其他实体进行了逻辑描述，它们中的绝大多数是在 OGSA 范围之外的，但它们引发了从支持使用案例的基础设施的角度来定义结构。

这三层需要互操作并协同工作以提供所需的服务质量，由这三层共同影响的服务质量被称为宏观质量。

2. OGSA框架

图 6-4 中，OGSA 根据 SOA（service-oriented architecture，面向服务架构）中的服务及其接口、服务所含资源的独立或集合状态，服务之间的交互实现了逻辑中间层。OGSA 服务框架如图 6-4 所示。下面是该框架的几个重要方面。

促进 OGSA 发展的重要动力之一是合成范式（compositon paradigm）或模块构建（building block）方法。这种方法按需要从原始的最小能力集中形成一组能力或函数来满足一个需求，而预先对这个需求并不了解。这提供了结构变化所需要的适应性、灵活性和健壮性。

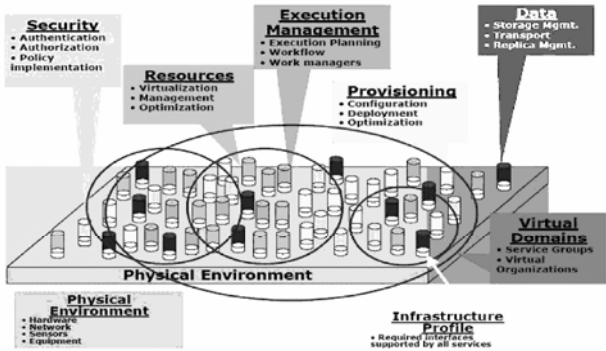


图 6-4 OGSA框架¹

¹ Global Grid Forum, The Open Grid Services Architecture, Version 1.0, [EB/OL].(2008-8-20), http://forge.gridforum.org/sf/go/projects.ogsa-wg/docman.root.published_documents.ogsa_1_0

OGSA 描述服务及其接口、语义/行为以及服务间的相互作用。需要注意的是，支持这些服务内部实现的软件结构不在 OGSA 工作组的讨论范围之内。

OGSA 结构不分层，一个服务的执行基于它逻辑上所依附的层而建立，并只与该层交互；或者说 OGSA 的结构是面向对象的——尽管许多概念似乎是基于对象的。

服务之间采取松散耦合，它们要么单独存在或作为一组相互作用的服务的组成部分，正是这种松散耦合通过服务之间的执行、合成、相互作用实现了 OGSA 的能力。例如，一方面，为了实现控制能力，一组服务可能被结构化以便组中的某些服务能够进行控制，而组中的其他服务提供被控制的接口和机制。一个服务可以参与多个服务集以实现不同的功能。但是，实现某个特定的能力不必调用组中的所有服务。

服务可能是实现某个能力的虚拟服务集的一部分，这个虚拟服务集称为虚拟域（virtual domains），如同一个服务组；或者共享一个集体环境或管理框架，如同出于一个虚拟组织中。

OGSA 服务要求并假定一个物理环境，该环境包括著名的物理元件和连接如计算硬件和网络，甚至是物理设备，如望远镜。

作为 OGSA 网络的服务必须有一组核心的非空接口、标准和常识/引导程序，它们被看作基础设施服务（infrastructure services）或网格结构（Grid fabric）。

图 6-5 显示了 OGSA 服务之间可能存在的大量关系及它们之间的相互作用。图 6-5 中柱状图代表了物理基础设施层之上的单个服务，服务建立在 WebService 标准之上，带有与网格相关的语义、附件、扩展和修正。图中的每一层都是一个集合，代表了一个功能实现——例如，一个层可能实现执行管理能力，另一个层可能实现数据管理能力。图中显示的关系是具体服务之间的关系，也可能跨越功能实现。服务间关系类型包括使用、组合、代理、参考、扩展等。服务之间为其他目的（如管理、服务的说明）而发生的相互作用也可以建立关系模型。

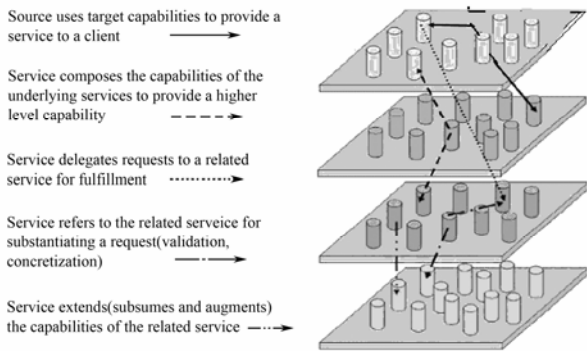


图 6-5 OGSA 服务之间的关系¹

¹ Global Grid Forum, The Open Grid Services Architecture, Version 1.0, [EB/OL].(2008-8-20), http://forge.gridforum.org/sf/go/projects.ogsa-wg/docman.root.published_documents.ogsa_1_0

3. 基础设施服务

定义 OGSA 的目的在于：在面向服务结构的环境中，定义一套紧密关联、集成整合的组件协同解决上面所提到的需求。因此，如果要详细描述上层服务，必须首先对其所依托的基础设施进行假定。有些规范指定的太过抽象，无法达到当初指定这些规范的目的。因此，指定 OGSA 规范就如同指定 TCP 和 DNS 等各种高层 TCP 协议与服务一样，这些高层协议需要底层属性来进行通信，那么 OGSA 的设计受到所选用的底层机制的影响。

最重要的假设是 OGSA 依托于正在形成并不断发展的 WebServices Architecture (WSA) 技术规范群，实际上，OGSA 被看作是核心 WS 标准的一个特殊应用框架。这一选择是基于对面向服务架构的优点信任，相信其是对网格系统所需功能做出广泛采用的、具有工业标准的面向服务的描述的最有效路线。

选择 Web Service 作为基础设施和框架意味着 OGSA 系统和应用是按照面向服务架构原则结构化的，服务接口用 WSDL 定义，并采用 XML 作为描述和表示语言，采用 SOAP 作为 OGSA 服务的主要消息交换格式，通过 WS Interoperability (WS-I) 过程制定与互操作框架定义一致的服务定义。

但是，Web service 标准并不能完全满足网格的需求。在某些环境下，需要对现有规范进行必要的修订或扩展。其中，对现有规范扩展的一个典型例子是扩展安全功能。OGSA 基础设施的各个层面都有安全问题，使用 WS 安全标准协议允许 OGSA 服务请求为授权认证及消息保护携带合适的令牌。OGSA 基础设施的一些环境中需要端到端的消息保护，因此，OGSA 必须提供更高层的保护机制如 XML 加密、数字签名以及点对点的传输层安全如 TLS 和 IPsec。除消息层的安全外，一个互操作的和组合的基础设施还需要安全组件本身作为服务来提供。目前，研究者们正在进行各种努力对这些安全服务进行定义，例如，OGSA 授权服务可能使用推荐的 WS 协议标准结合正在发展的 OASIS 标准（包括 SAML & XACML）表达安全状态和存取控制。

4. 执行管理服务 (Execution Management Services, EMS)

执行管理服务 (OGSA-EMS) 主要管理和表示完成单元任务。这里的单元任务包括 OGSA 应用和传统应用（非 OGSA 应用如数据库服务器）。EMS 的概念元模型如图 (6-6) 所示。

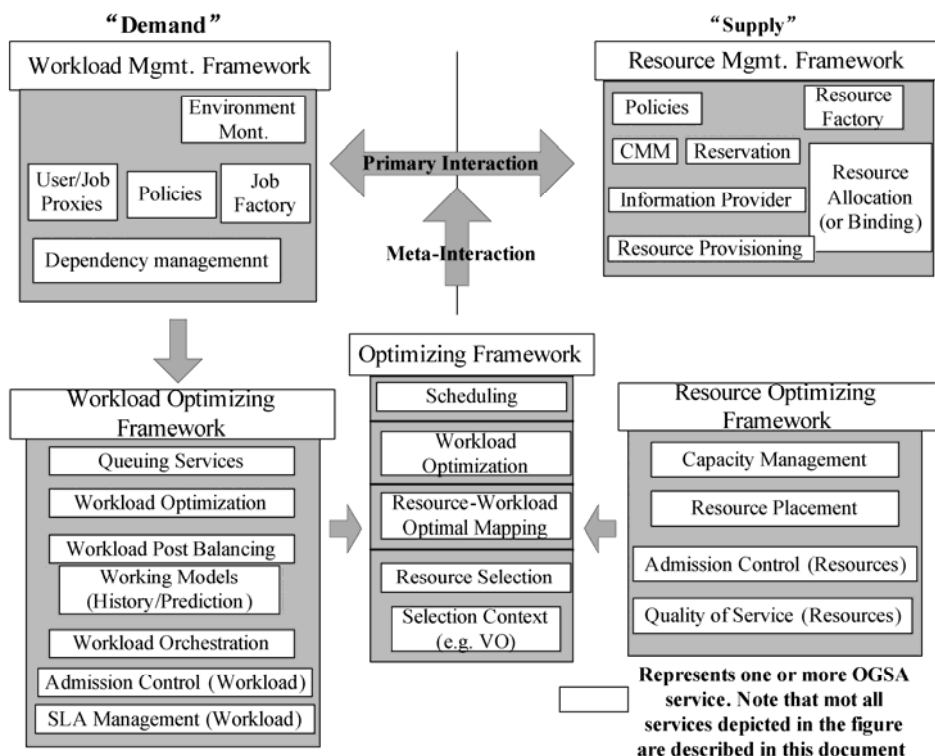


图 6-6 EMS 的概念元模型（分解为供应方和需求方）¹

这里，EMS 解决的问题可以用下面的例子来说明：一个应用需要一个缓冲服务，那么应该使用现有的服务还是创建一个新的服务；如果创建了一个新的服务，这个服务应该放在什么位置；如何为它分配足够的资源（CPU、内存、磁盘等）；缓冲服务应该使用何种服务协议；需要何种协议。

从上述例子中可以概括出，EMS 解决的单元任务执行中的问题包括任务的安排、供应和生命周期管理，包括但不限于以下内容：

① 发现执行候选位置（Finding execution candidate locations）：确定执行单元任务的最佳位置，这个位置应该提供任务执行所需的各种资源如 CPU、内存、可用许可等。应该设置何种限制策略来进一步限制执行候选位置的范围。

② 选择执行位置（Selecting execution location）：一旦知道了单元任务可以在何处执行，剩下的问题就是确定究竟要在何处执行。要解决这个问题，需要运用不同的选择算法，这些算法可以优化不同的目标函数、或尝试加强不同的策略与服务水

¹ Global Grid Forum, The Open Grid Services Architecture, Version 1.0, [EB/OL].(2008-8-20),

http://forge.gridforum.org/sf/go/projects.ogsa-wg/docman.root.published_documents.ogsa_1_0

平协议。

③ 准备执行 (Preparing for execution)：要真正执行一项单元任务，仅仅知道合适的执行位置还不够，还必须进行必要的配置，包括对二进制码、各种库、阶段性数据或其他形成本地执行环境所需要的操作进行部署和配置。

④ 初始执行 (Initiating the execution)：相关环节准备完备之后就可以开始执行任务了，同时也可以进行其他的相关工作，如在合适的位置注册该任务。

⑤ 管理执行 (Managing the execution)：任务启动之后，直至结束之前，必须对其进行管理和监测。如果任务失败该如何处理，任务未能达到其目标该如何处理？可否在另一个位置重新开始该任务，任务的状态如何，各方面都需要进行管理和监测。

上述问题是 EMS 解决的主要问题，它涉及任务完成的各个阶段、各个方面。EMS 非常重要，因为在不同环境中应用网格，其中的资源类型及对资源的使用是千差万别的，而共同点是监测应用需求并根据需求做出动态的反应来完成任务。

这一功能是通过一组服务实现的。这些服务将 EMS 问题分解成多种可替代的元件，在不同的实例中可能用到不同的服务组合来实现它们的目标。一般来说，EMS 由供应方和需求方组成，供应方提供资源如 CPU、磁盘、数据、内存、服务，需求方使用这些资源。需求方有一系列工具如工作负载管理工具、工作流系统等，而供应方有管理和供应资源的服务。

EMS 包括以下三种类型的服务

- ① 资源 (Resources)：模拟过程、存储、执行、资源管理和供应；
- ② 任务管理与监测服务；
- ③ 资源选择服务：集中决定在何处执行单元任务。

5. 数据服务 (Data Services)

OGSA 的数据服务实现了数据资源的存取、管理和更新。数据服务旨在将数据移动到需要它的位置，管理数据的复制，执行查询和更新操作，将数据转换成新的格式，同时提供对描述 OGSA 数据及其服务的元数据的管理。

数据服务中的数据资源指可以作为数据来源或沉积位置的实体。网格的异构性意味着必须能够支持各种不同类型的资源，包括普通文档、数据流、数据库管理系统、目录、衍生数据以及数据服务本身。

数据服务的使用环境包括远程存取、分段传输 (Staging)、复制、数据联合 (Federation)、数据衍生 (Derivation，从一种数据资源自动生成新的数据资源)、元数据管理。

数据服务的供需概念元模型如图 6-7 所示。

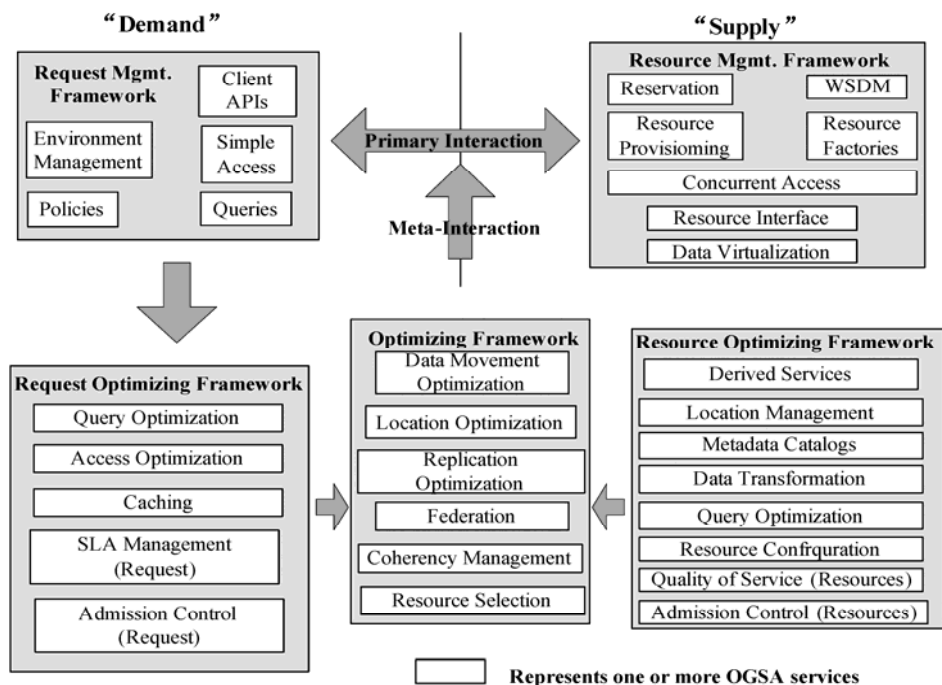


图 6-7 数据服务的供需概念元模型¹

数据服务的能力包括：透明和虚拟化、支持客户端 APIs、支持可扩展的数据类型及其操作、数据位置管理、简单存取、查询（结构化存取）、数据转换、数据更新、安全映射扩展、数据资源配置、元数据、计量、服务质量、性能、一致性、可用性等。

6. 资源管理服务（Resource Management Services）

资源管理是对网络资源所实施的几种管理，OGSA 网络上有三种与资源有关的管理：

- 对资源自身的管理（如重启主机、在网络上设置 VLANs）；
- 网络上资源的管理（如资源存储、监测和控制）；
- OGSA 基础设施的管理（如注册服务的管理）。

在资源层，直接通过本地管理接口管理资源。这一层面的管理包括监测、配置和控制以及发现。

基础设施层提供了资源的基本管理行为，形成了 OGSA 环境中的管理基础。为

¹ Global Grid Forum, The Open Grid Services Architecture, Version 1.0, [EB/OL].(2008-8-20),

http://forge.gridforum.org/sf/go/projects.ogsa-wg/docman.root.published_documents.ogsa_1_0

了整合大量各种类型的资源，需要对这些基本管理行为进行标准化操作。基础设施层提供了一个基本的管理模式和一个通用的管理接口。

在 OGSA 网格中，不同形式的管理由不同类型的接口实现，可以把这些接口归纳为三个层次（参见表 6-1）。

表 6-1 管理类型与接口之间的关系

管 理 类 型	接 口 层 次	接 口
资源本身的管理	资源层	CIM/WBEM, SNMP 等
	基础设施层	WSRF, WSDM 等
网格上的资源管理	OGSA 功能层	功能/函数接口
OGSA 基础设施的管理		特定管理接口

在 OGSA 功能层，有两种管理接口：

- 功能接口（Functional Interface）：OGSA 的一些功能（如 OGSA EMS）是以一种资源管理形式表现出来的，这些能力的服务通过提供功能接口来实现（如建立或结束一项任务）。
- 管理接口（Manageability Interface）：每一项功能都有一个特定的接口，通过这个接口可以管理这项能力。

7. 安全服务（Security Services）

OGSA 安全服务旨在推动组织/虚拟组织加强与安全相关的策略。加强安全策略的目的是确保满足更高水平的商业目标。加强策略意味着成本的增加，由于执行不严格会产生损失，但策略越严格，对利润的负面影响就会越大。这些因素之间存在着复杂的、微妙的关系。

网格应用可能跨越多个管理实体，不同的组织有自己的商业目标，并按照各自的目标建立并加强自己的策略，这些策略在复杂性和严格程度上存在很大差别。为了满足商业或安全目标，安全策略在如下方面做了规定：消息完整性和保密性的加强、相互作用的实体的授权、最小化授权进度、安全记录日志和审计、最低的特权操作和入侵监测等。

OGSA 安全服务的构成元件必须以一种能够实现各个系统安全互操作的方式来支持、整合和统一各种流行的安全模型、机制、协议、平台和技术，组件必须支持与现有安全结构和模型在跨平台、跨主机环境下进行整合。

总的来说，实体在特定环境中通过一定的机制相互作用，其中，实体包括用户、服务、主题等事物，机制包括不同交流方式的组合，如电子邮件、电话、HTTP、SOAP 等。所有这些实体、机制和环境能够被一组属性所描述，其中一些属性的类型或值可以用作唯一识别，其他的则用于分类和分组。安全策略是关于这些不同的实体、

相互作用的机制和环境的声明，以及指定相关属性、属性值及其相互关系的限制。策略声明将按照实体、资源、环境特征加以表示。这些策略将涉及包括授权认证、信任、身份认证、代理等方面。

8. 自管理服务 (Self-Management Services)

自我管理被看作是帮助降低成本、减少拥有和操作信息技术基础设施复杂性的一种方式。在自我管理环境中，系统组件包括硬件设备如计算机、网络、存储设备和软件如操作系统下面重点讲述自管理的有关情况。

自我管理的主要目标之一是使一组服务达到一定的水平，能够尽量实现自治，降低成本和系统管理的复杂性。通常情况下，在操作环境中，以一种组件开发者不能事先确定的方式控制一个解决方案构成组件的行为的各个方面是十分必要的。在一个自我管理系统中，可以通过部署各项策略来控制系统组件的行为实现上述要求。其中服务层次管理器 (Service Level Manager, SLM) 负责部署和调整策略，然后依据在系统中观察到的条件改变被管理资源及服务的行为，以确保与商业目标的整体兼容性。因此，必要时需要对不同的 SLM 进行组合和分层以降低系统初始设计和操作的复杂性。尽管自我管理是 OGSA 的一个重要的组成部分，但这项工作仍处于准备阶段，因此这里仅仅描述了自管理的某几个方面。

自我管理的各个阶段所涉及的属性包括：

(1) 服务层次协议 (Service Level Agreements, SLA)：包括服务提供者和服务使用者之间的商业和技术协议。SLA 为有关服务的目的、达到的目标提供指南。

(2) 策略 (Policy)：用于控制 SLM 的行为及其可管理的资源。控制自我管理实体的策略摘自 SLAs，并且作为管理资源使用环境的一部分。

(3) 服务层次管理模型 (Service Level Manager Model)：提供演示 SLM 并与合作的框架，以便各种 SLMs 和操作人员能够相互理解，即便他们不了解设计时的相关知识。

自我管理的典型应用环境包括两个层次，第一层是任务层次的管理。IT 商业活动管理者提交一项任务并协商任务的 SLA，任务必须执行以满足特定的协议。在稍后的阶段，IT 商业活动管理者可能希望重新协商协议以满足新的商业需求，这将通过任务管理者来完成并更新现有的协议，然后在服务级别重新计算资源需求并触发供应步骤的执行。IT 商业活动管理者能够使用监测服务和计量服务来监测负载和资源的利用情况，同时也可以从任务管理器获得正在执行的任务的状态。第二层是网络系统层次的管理。在网络系统层次管理的目标之一是改善资源的利用情况，维护执行任务的 SLA。网络系统层次管理者需要为预计的负载增加而添加新的资源，为降低成本而释放多余的资源，分配给任务的资源也会根据策略进行调整，如在一些任务优先于另外一些任务的情况下。

实现自我管理的机制包括自配置机制和自愈合机制：自配置机制指的是利用 IT 专

家提供的策略进行动态调整以适应 IT 系统的变化；自愈合机制指监测资源和服务中不合适的操作，在不中断 IT 环境的情况下启动基于策略的校正行动自愈合，如自我保护，这使得组件可以检测到敌对的行为（如非授权的存取和使用、病毒等），采取正确的行动，优化自身机制，进行自动调整，以最高的效率满足终端用户的需求或商业需求。这种调整可能意味着重新分配资源以改善整体利用效率，或者改进 SLA。

9. 信息服务（Information Services）

在网络环境中，高效的存取和操作相关应用、资源和服务的信息是 OGSA 的一个重要的能力。这里的信息指用于状态监测的动态数据和事件，用于发现相关的静态数据，以及日志记录下来的任何数据。

OGSA 信息服务的用户包括（但不限于）：执行管理服务、记账服务、问题确定服务、资源预约服务、资源使用服务、应用监测。为了促进互操作和重用，信息服务本身应该建立在 OGSA 的基础设施能力之上，同时，应该充分利用其他的 OGSA 服务。

对信息服务的描述在很大程度上依赖于信息来源、目标及服务质量的需求。信息服务的常见结构是：信息生产者希望向多个信息消费者提供信息，信息消费者希望从多个信息生产者处获得信息，消费者可以直接联系生产者，也可以利用订阅机制来获取信息。OGSA 并未描述用于实现信息服务的数据模型，也未描述查询信息的语言。当前系统多基于 XML、XPath/XQuery 查询语言或使用关系模型或 SQL 查询语言。

OGSA 有多项信息服务，一些是通用的，另一些被加以优化以满足特定的用户实例。信息服务可能允许生产者和消费者提供自身相关的详细信息以供查询，使得彼此发现对方。使用特定的分布式注册机制或点对点机制可以实现这一目标。详细的描述可以包括诸如生产者或消费者的类型，他们所生产或消费的信息，以及这些信息的 URL 等信息。

下面是几个信息服务的例子。

（1）目录服务：通过目录的形式提供信息供用户检索。OGSA 的目录服务基于 WSRF 服务组概念。

（2）日志记录服务：大量分布式计算使用环境中需要用到日志服务。这些环境包括问题确定、使用计量、错误恢复、事务处理以及安全。

（3）在生产者/中介/消费者模型中完成向消费者提供信息。在分布式计算环境中，传统的生产者/中介/消费者模型中的一些服务由信息服务来完成，如名称服务、日志记录服务、通知服务等。

6.2.2 开放网格基础框架

OGSA 自身无法单独准确地描述网格服务的细节，它仅仅给网格服务描绘了一个轮廓。这也是为什么 OGSA 又创造了 OGS I 标准来规范地描述网格服务技术细节的原因。它们之间的区别如图 6-8 所示。

OGSA、OGSI 以及 GT3 可以通过一个例子来说明。假如现在要盖一所房子，要做的第一件事情是雇佣一个建筑师来画出蓝图，这样才能知道要建造的房子大概是什么样子。接下来需要雇佣一个工程师来计划一下建筑细节，比如哪里放大梁，哪里走电缆等。工程师把他所有的计划告诉合格的专业工人们，比如建筑工人，电器工人等。最后由工人来完成施工。

在这里，OGSA（负责概念定义）就是上面例子中的建筑师，OGSI（负责详细规范）是工程师，而 GT3 就是在工程师的详细规范中使用到的砖头、水泥、大梁等。GT3 提供了一个完整的开放网格服务基础设施应用。GT3 将服务发现、程序执行作业的提交、监控和可靠的文件传输等功能定义成与 OGSI 兼容的服务。其他如数据传递、副本定位和授权等服务也尽量与 OGSI 相兼容，并且 GT3 定义了一组关于使用 Web 服务描述语言（WSDL）和扩展标识语言（XML）模式的约定与扩展，以便启用有状态服务。

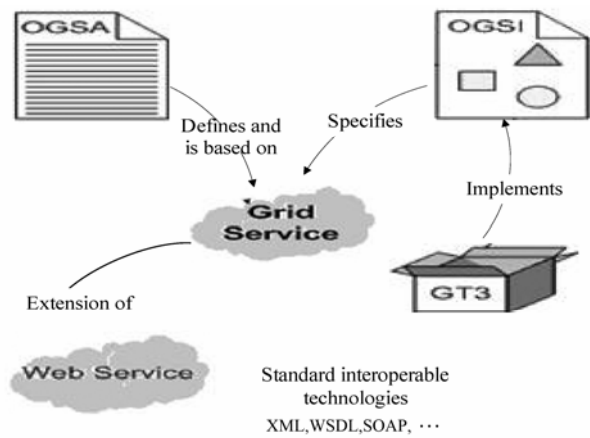


图 6-8 OGSA、OGSI以及GT3 在网格服务架构以及Web服务中的地位¹

¹ OGSA, OGSI, and GT3, [EB/OL].(2008-8-20),<http://docs.huihoo.com/globus/gt3-tutorial/ch01s01.html>

但 OGSI 过分强调网格服务和 Web 服务的差别,导致了两者之间不能更好地融合。由于 OGSI 单个规范中的内容太多,所有接口和操作都与服务数据有关,缺乏通用性,而且 OGSI 规范没有对资源和服务进行区分。OGSI 过度采用 XML 模式,比如 xsd: any 基本用法、属性等,其这种性质可能造成移植性差的问题。这使得 OGSI 无法使用目前的 Web 服务和 XML 工具实现良好工作。

6.2.3 数据存取与集成标准 (OGSI-DAI)

OGSA-DAI 是开放网格服务架构数据存取和集成的网格标准,在 Globus Toolkit 上进行开发,支持 DB2、Oracle、Xindiee、MySql 等数据库管理系统。网格数据库是对现有数据库的网格化,基于开放网格服务体系结构提供网格数据库服务,使网格用户或其他网格服务可通过网格数据库服务访问网格中的各种异构数据库,从而达到数据资源的高度共享和协同处理,对数据资源的访问更加透明、高效、可靠,网格数据处理的能力更强,可以满足虚拟组织的数据处理需求。

OGSA-DAI 项目致力于建造通过网格访问和集成来自不同的孤立数据源的中间件。这个项目由 UK Database Task Force 提出构想,并紧密地和全球网格论坛数据访问和集成服务工作组 (GGF-DAIS WG) 以及 Globus 团队一起工作。总体上,OGSA-DAIS (Open Grid Service Architecture-Data Access Integration Service) 与 DAIS 相符合,也努力成为 DAIS 网格数据库服务推荐标准的第一个参考实现。

OGSA-DAI 的目标是通过网格进行数据访问和集成,并提供统一的服务接口。通过 OGSA-DAI 的接口,不同的异构的数据源和资源被视为逻辑上的单一资源,同时这些资源可以在 OGSA 框架内集成。OGSA-DAI 网格服务提供基本的操作来完成复杂的操作,比如数据联盟、在虚拟组织进行分布式查询等,但是它隐藏了如数据库驱动、数据格式和客户端的传输机制等技术细节。

DAIS 工作组成立于在 2002 年 2 月召开的 GGF4 会议,它旨在寻求促进与 OGSA 相适应的网格数据库服务的标准,初衷是提供对现有的、自主管理的数据库的一致访问,而不必开发一个新的数据存储系统。更准确的说,是要在网格框架内更易于单独的或共同的使用这些系统。DAIS 将于 GGF9 上提出一个网格数据库服务的初始版,今后也将支持更广泛定义的数据库的访问和集成,例如,文件系统、来自仪器和设备的数据流等。

目前使用 OGSA-DAI 的项目有: AstroGrid、Biogrid、BioSimGrid、Bridges、FirstDIG、GeneGrid、ODD-Genes、OGSA-WebDB。

OGSA-DAI 是科学方法集成的关键。它能够集成研究成果和出版物,观测、仿真和实验的数据,促进对数据的新的应用;对方法和结果进行校验。通过集成各自收集资源,促进新的科学发现。OGSA-DAI 同时是大规模协作的关键。OGSA-DAI 在 OGSA 中的位置如图 6-9 所示。

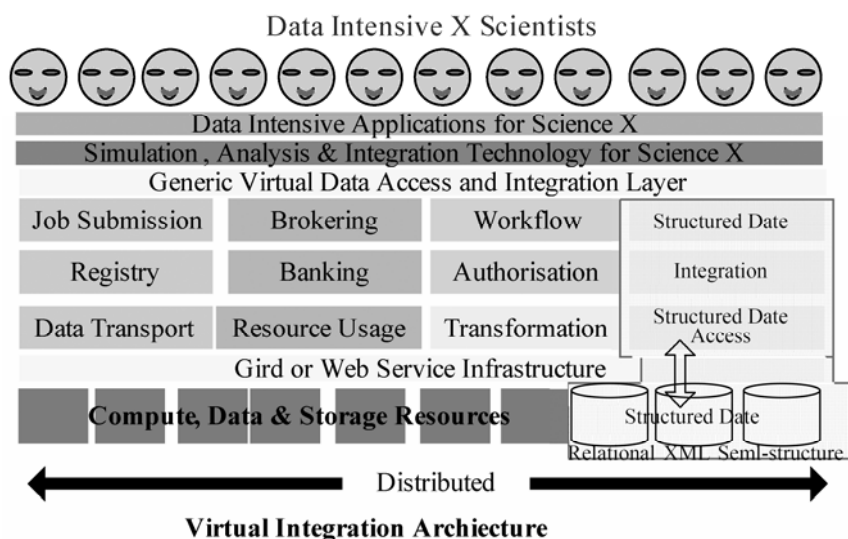


图 6-9 OGSA-DAI在网格中的位置¹

OGSA-DAI 的体系结构与 WebService 相似，如图 6-10 所示。



图 6-10 OGSA-DAI体系结构²

从图 6-10 中可以看出，其体系结构与 WebServices 很相似，都是 Discover、Bind、Publish 机制。

OGSA-DAI 主要有三种服务类型（见图 6-11）：

（1）服务组注册器（Service Group Registry，SGR），通过服务组注册器可以找到所需要的网格数据服务（Grid Data Service，GDS），也可以通过它找到所需的网格数据服务工厂（Grid Data Service Factory，GDSF）。

¹ Malcolm Atkinson, An Introduction to Data Services & OGSA-DAI, [EB/OL].(2008-8-20), <http://www.dma.unina.it/~murli/GridSummerSchool2004/session-27.htm>

² Malcolm Atkinson, An Introduction to Data Services & OGSA-DAI, [EB/OL].(2008-8-20), <http://www.dma.unina.it/~murli/GridSummerSchool2004/session-27.htm>

- (2) 网格数据服务工厂，这项服务用于创建一个 GDS 实例，以访问特定的数据资源。
- (3) 网格数据服务，通过这项服务可以访问某个数据资源（关系数据库或 XML 数据库，甚至是存储在普通文件中的数据）。

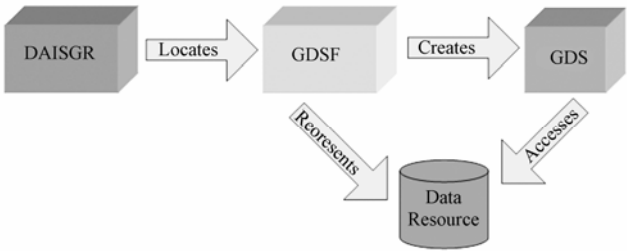


图 6-11 OGSA-DAI的服务及主要功能¹

6.2.4 Web服务资源框架

为了解决 OGSi 和 Web 服务之间存在的矛盾，2004 年 1 月，HP、IBM、富士通以及 Globus 提出了一系列的详细规范来解决这个问题，即 Web 服务资源框架（WebServices Resource Framework，WSRF）以及 Web 服务通知（WebServices Notifications，WSN）。图 6-12 是有关详细规范的发展路标图。

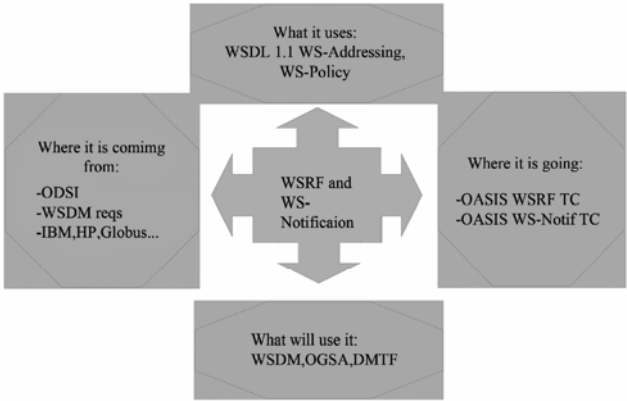


图 6-12 WSRF 与 WS-Notification 发展路标图

WS-Resource Framework（WSRF）是 6 个关于 Web 服务的详细规范的定义集合，这些详细规范定义了 Web 服务上下文中的 WS-Resource 方法、建模以及管理状态。WSRF 是一组规范，提供一种标准方式，在 Web 服务应用程序的本质“无状态的”环境中与

¹ Malcolm Atkinson, An Introduction to Data Services & OGSA-DAI, [EB/OL].(2008-8-20), <http://www.dma.unina.it/~murli/GridSummerSchool2004/session-27.htm>

“有状态”资源交互。这种有状态资源可以是任何东西，从数据库到电子鼠标都是有状态资源。事实上，可以使用 WSRF 来处理任何可以通过改变其属性来操纵的对象。这里，WSRF 采用了与网格服务完全不同的定义：资源是有状态的，服务是无状态的。为了充分兼容现有的 Web 服务，WSRF 使用 WSDL 1.1 定义 OGSi 中的各项能力，避免对扩展工具的要求，原有的网格服务已经演变成了 Web 服务和资源文档两部分。

WSRF 试图为 Web 服务所面对的最常用的涉及有状态资源交互活动的过程提供一种访问有状态资源和消息交换模式标注重集，其主要限制是：模式和消息传递交换必须符合 Web 服务架构；这个详细规范必须能够和 Web 服务其他已经成为标准或者将要成为标准的其他规范共同工作。

WS-Resource 是一个有状态资源（如数据库或硬盘）和它与之交互的 Web 服务的组合。WSRF 作为一种规范，定义了描述这些操作的 WSDL 文件的结构。该 WSDL 文件可以被任何语言所解析。WSRF 规范只定义应该做什么，而不定义应该如何做。这里描述的概念的实际实现留给应用程序去完成。

WSRF 基本保留了 OGSi 中的所有功能，同时更改了一些语法，并且还在其表示中采用了不同的技术。Web 服务通知（WSN）为 Web 服务提供基于消息发布和预定的能力。WSRF 和 WSN 都是建立在已存在的 web 服务定义和技术基础上，帮助实现了网格计算、系统管理和 Web 服务的统一。

2005 年初发布的 Globus Toolkit 4（GT4），实现了 WSRF 和 WSN 标准。GT4 提供 API 来构建有状态的 Web 服务，其目标是建立分布式异构计算环境。所有知名的 GT3 协议都被重新设计为可以使用 WSRF，并且 GT4 也在其中增添了一些新的 Web 服务的组件。

Web 服务包括无状态的 Web 服务以及有状态的 Web 服务两类，其关系如图 6-13 所示。

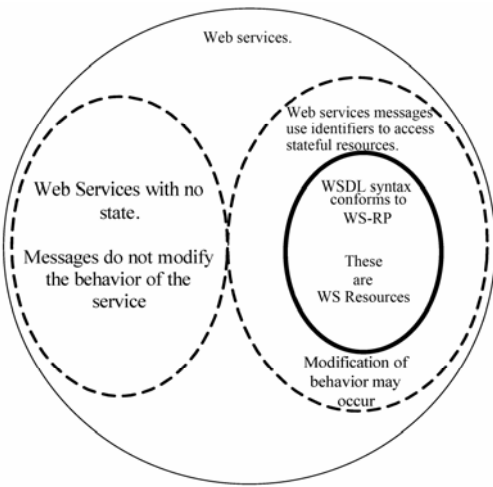


图 6-13 Web 服务与 WS Resources 和无状态 Web 服务的关系图

如图 6-14 所示，WSRF 的目标是将其整合进入 Web 服务标准的家族中，而不仅仅是一个像 OGSI 一样的对 Web 服务的补丁。在这个情景下，OGSA 将直接基于 Web 服务而不需要基于 OGSI 网络服务。

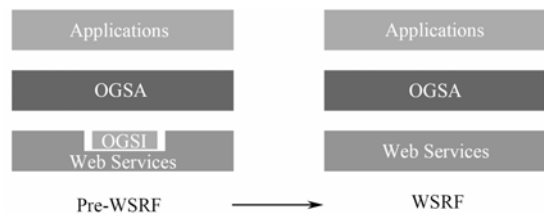


图 6-14 WSRF 出现的前后架构变化比较

WSRF 新规范并没有很大地影响 OGSA。所有在 OGSA 中定义的高层服务都将保持相同的接口和详细规范，比如工作管理，资源管理，安全服务等。唯一不同的是，下层的中间件将是纯粹的 Web 服务而不是 OGSI 网络服务。

WSRF 到底在多大程度上真正支持了 Web 服务，实际上，它能够克服 OGSI 的缺点：

- （1）OGSI 详细规范冗长并且密集，而 WSRF 详细规范被分为了五个文档，加上一个补充规范（WS-Notification，可以不严格算作 WSRF 的一部分）。
- （2）OGSI 不能和目前的 Web 服务工具很好地协调工作。WSRF 非常重视目前绝大多数的 Web 服务社区中的缺陷，这使得目前的 Web 服务工具将会很方便地获得 WSRF 的支持。
- （3）OGSI 过于面向对象。WSRF 清晰地将状态与服务分开，纯粹的 Web 服务不再有状态，对 factory/实例服务的支持也消失了，但是仍然能够通过使用设计模式来实现。

6.3 与网格相关的共享和集成技术

除了网格本身的标准和技术之外，还有很多信息处理技术为网格系统提供支撑和支持。这些技术主要有 Web 服务技术、Ontology 和语义（门户）网格技术、网格门户技术。

6.3.1 Web服务技术

在 e-Science 中，Web 服务技术主要涉及 SOA、WebServices、WS-I 和 WS-I+。

1. SOA

2002 年 12 月，Gartner 提出“面向服务的架构（SOA）是现代应用开发领域最重要的课题”之后，国内外计算机专家、学者掀起了对 SOA 的积极研究与探索热潮。

SOA 不是一种具体的实现技术，它是一个抽象的概念和指导方针，可以用任何一种现今成熟的技术实现它，也就是说 SOA 是一种构件模型。

简单而言，在传统的方式下一个大型应用是由不同功能模块构成。而在 SOA 下，这些功能模块将被功能服务的概念所取代，这些所谓的功能服务通过定义良好的接口和协议联系起来。而这些接口的定义是采用中立的方式进行的。它独立于实现这些服务的硬件平台、操作系统、网络架构和编程语言。在一个大系统中各种各样的服务，甚至是不同软件供应商所提供的其他应用都可以按一种统一、通用的方式进行交互。由于 SOA 只是一种构件模型，必须用一些现有的技术去实现它，如 DOCM（Distributed Component Object Model）、CORBA（Common Object Request Broker Architecture）、Web Service 等。

由于接口是中立的，SOA 还牵涉到另一个概念“松散耦合”，耦合度越低系统开发就越灵活，更容易维护。而面向服务的构架就是一种“松散耦合”的系统架构，把原来在代码中的功能提到应用服务的层面。

SOA 的关键是“服务”的概念。Web 服务是实现 SOA 最好的方式。但 Web 服务本身还有很多不成熟的方面，在可靠消息传递、安全 Web 服务、Web 事务处理等方面的标准还有待完善。

2. Web Services

WebService 是基于网络的、分布式的模块化组件，它执行特定的任务，遵守具体的技术规范，这些规范使得 Web Service 能与其他兼容的组件进行互操作。它可以使用标准的互联网协议，如超文本传输协议 HTTP 和 XML，将功能体现在互联网和企业内部网上。Web Service 平台是一套标准，它定义了应用程序如何在 Web 上实现互操作性。Web Service 提供永久的无状态服务，实现了程序间远程访问的透明性。

Web Service 技术实现了 Web Service 的定位、访问与管理，是互联网工业的事实标准。有关组织已经发布了一些实现万维网服务所需要的协议和标准，如 SOAP、WSDL、UDDI 等，详细框架见图 6-15。

很多 OGSA 背后的概念都是建立在 Web Service 标准之上的。从很多方面来讲，网络操作就像 Web Service 一样，都是由一个系统向另一个系统提出执行某项任务的请求，并获得一个响应。然而，Web Service 有很多与生俱来的特征，使其不适合在网络环境中使用。首先，Web Service 是无状态的，它们不会记住从调用到调用过程中用户请求（或收到）了什么。如果用户要进行一系列相关的计算或查询，就必须找到一种方法将一次计算的结果传递给下一次调用。OGSA 采用了万维网服务的 WSDL 和 SOAP 规范。遵守 OGSA 标准的系统都可以连在一起，可以很容易地进行集成，共享各种系统提供的功能。

3. WS-I和WS-I+集：Web服务详细规范

WS-I 代表 Web 服务互操作组织（Web Service Interoperability Organization），它

是一个开放的产业促进组织，目标在于促进 **Web Service** 在不同的平台、应用软件和编程语言中实现互操作，它同时代表一个标准集成者，帮助 **Web Service** 朝结构化的、协调化的方向迈进。

目前 **WS-I** 的工作重点在于描述、发现、传输、绑定和安全。如图 6-16 所示。

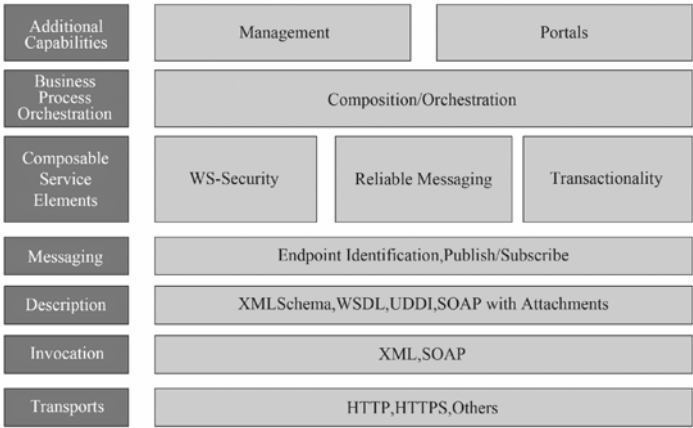


图 6-15 Web服务框架¹

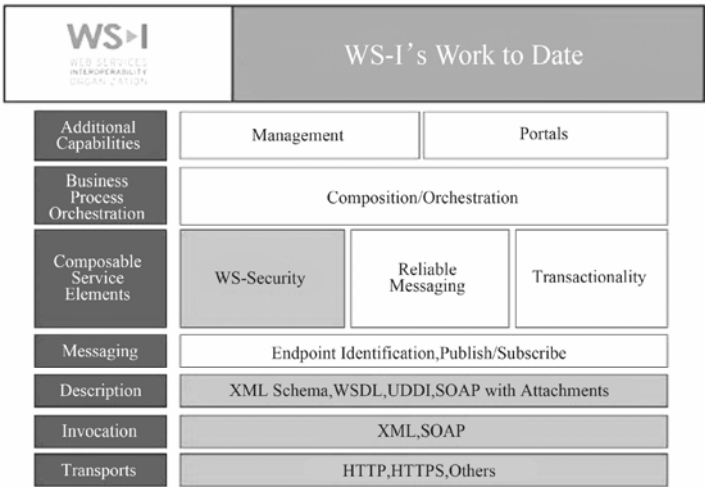


图 6-16 WS-I²

¹ David Burdett, Interoperability: Ensuring Success of Web Services, [EB/OL]. (2008-8-20), http://www.ws-i.org/docs/presentations/wsi_cidx_preso_110105.ppt

² Andy Astor, Interoperability: Ensuring the Success of Web Services, [EB/OL]. (2008-8-20), http://www.ws-i.org/docs/presentations/wsi_idg japan_preso_final.ppt

OMII (Open Middleware Infrastructure Institute) 在 WS-I 的基之上, 定义了一套包括 WS-I, plus BPEL, WS-RM 和 WS-Addressing 的针对 e-Science 和网络的 Web Service 详细规范, 被称为 WS-I+。

OMII 的 WS-I+ 建议是在已有的 WS-I 的基础上添加了一些重要的详细规范。WS-I 规范已经提供了 XSD (XML Schema Definition)、WSDL、SOAP、UDDI 以及部门 WS 安全内容, WS-I+ 中的规范还包括服务发现、工作流、消息、寻址和通信等。

核心服务架构: XSD、WSDL 和 SOAP。XSD、WSDL1.1 以及 SOAP1.1 是 WS-I 内容的一部分, 并且将会被 OMII 使用来提供基于 Web 服务的网格的基础框架。这些详细规范的新版本, 比如 WSDL2.0, 也将会在其得到认可的时候被使用。

服务发现: UDDI。因为 UDDI 是 WS-I 内容的一部分, 所以它可能会被 OMII 作为 WS-I+ 的一部分继续使用以构建注册基础框架。然而, 现有的项目已经利用 WSI 提供的 Web 服务构建了它们自己的注册部分, 到目前为止, 还不清楚 UDDI 详细规范能否为 e-Science 的应用提供有效的帮助。未来, OMII 可能会通过扩展的方式补充核心的注册服务, 尤其是那些基于语义网格的方法, 因为它们在服务寻找或者扩展了欧洲数据网格项目 (EDG) 的 R-GMA 的解决方案添加了更进一步的语义信息。

工作流: 或者说“为网格制定规划”, 被认为是 e-Science 的一项非常重要的功能, 这方面的近期工作在 GGF10 workshop 中有所总结。工作流是将服务联系起来的常规方式, 这里可能存在几条标准。既然 BPEL 被 IBM、微软、Oracle 以及其他公司所支持, 将这个标准包含在了 WS-I+ 集中。然而, 由于 BPEL 并非专门应用于 e-Science 应用项目, 那还需要探索它在使用中的适用性以及可靠性, 将来可能会使用它扩展支持高性能的数据流传输。

消息: Web Service 可靠消息。WS-ReliableMessaging (WS-RM) 和 WS-Reliability 几乎是相同的概念, 它们使用了消息队列和消息确认来保证消息在流中传递的可靠性。然而, 两条建议在这种情形下的联合也是可行的, 因为它们在架构方面达成共识, 而仅仅在语法的细节上有所差异。所以, 一个过滤器将能够直接在两条标准之间构成映射。当可靠性成为非常重要且苛刻的需要时, 将在 WS-I 中采用 WS-RM, 并且提供一个简单的工具, 以便在必要情况下向 WS-Reliability 映射。

寻址: WS-Addressing。WS-Addressing 是由 IBM 和微软提议的一条简单详细规范, 这条规范虚拟了寻址过程, 而它也被其他详细规范所推荐, 包括 WSRF。该规范允许独立定义传输协议中的 “end-point references”。这种能力对于构建网格是非常重要的。然而, WS-Addressing 目前还没有被提交给一个标准或者一个完整的协议。Oracle、Sun 和其他公司的 WS-MessageDelivery, 已经被提交到 W3C, 并且提供了比 WS-Addressing 更为丰富的特性集。这些特性集包括了有趣的 “抽象消息传输属性” 概念, 这个概念被定义在一个很宽泛的上下文中, 甚至包括非 SOAP 的传输。尽管目前 WS-Addressing 还尚未成为一条正式标准, 它还是被使用在 WSRF 中, 相信在 WS-I 中采用它将会对未来的 Globus GT4 版本及其他基于 WSRF 的解决方案的整合

起到很大帮助。然而，对于 WS-MessageDelivery 的丰富特性的学习还是很有用的，尤其考虑到 e-Science 需要 Web Service 与用户之间传输的高性能标准。

通知：通知覆盖了用来提供状态警告的消息，用户经常通过签署或者监听此类的“中断”实现通知。这种交互范例成为基于事件的编程模型的基础。Java 消息服务（Java Message Service, JMS）是一种可以通过一个良好的代理实现点对点（服务对服务）的操作的很好的解决方案。WS-Eventing 是一个来自微软的简单的点对点标准，同时，WSN 通知协议是由 IBM 和其他公司提出的一系列提供了更加全面的包括代理和点对点解决方案的服务集合的详细规范。

4. Ontologies技术

在哲学领域，Ontology 的意思是存在论。在知识共享的今天，在 e-Science 框架内，首先可以肯定的是，Ontology 这个术语和它在哲学上的含义是不同的。在 e-Science 领域，Ontology 表示的是一种概念化的详细规范。也就是说，一个 Ontology 其实是一个可以作为一个代理或者作为一系列代理而存在的概念及其之间关系的描述，比如对于一个 e-Science 项目的正式的详细规范。这个定义与 Ontology 作为概念定义集合的作用是一致的，但是要更加宽泛一些。

在 e-Science 范畴内，针对科学试验数据所构建的 Ontology 一般需要达到以下几点：

- ① 支持试验中点对点的数据以及元数据采集和共享。
- ② 支持资源处的数据发布。包括：制作可以出版的实验报告，自动的自然语言生成等。
- ③ 支持可以访问的在线出版物。
- ④ 对收集的数据和记录进行加工，并将其整理完备。
- ⑤ 使加工好的数据可以提供给人和其他计算机使用。

从图 6-17 中可以了解构建一个 Ontology 的框架示意图。

虚线内是试验数据的生成部分，其中包括科学试验的规划，工作台的相关事宜，试验数据的评估和测量以及计算服务。试验数据在产生之后一共有两条输出路线。

第一条，外部的路线代表产生的结果通过一系列相关服务的处理，生成结果数据，结果数据再由可视化工具展现给用户。

第二条，内部的路线代表产生的结果通过基于 SOAP 的 RDF 格式被保存成语义数据，之后，这些语义数据被语义 Web 应用所调用。

科学试验规划部分通常分为以下几个组成部分：制定科学试验计划；执行科学试验；思考科学试验中的问题；发布科学试验中的数据与资料。除此之外特别需要注意的是，在制定实验计划的时候，需要考虑以下两个问题：第一是实验计划工具的选择；第二个问题是实验需要基于某种规范来进行，比如 COSHH（Control of Substances Hazardous to Health）格式。

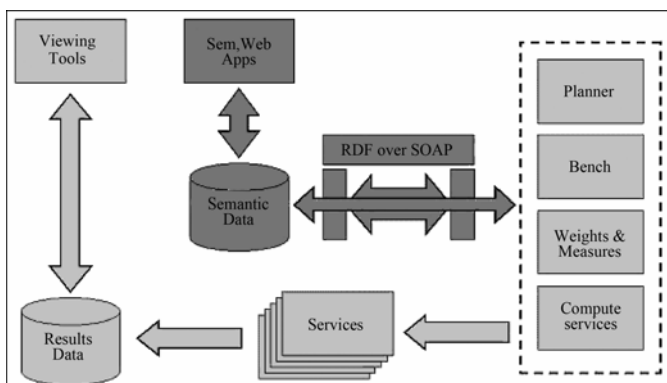


图 6-17 Ontology构建的框架示意图¹

在设计 Ontology 构建框架的过程中，需要考虑的问题有：

- (1) 在两个终端都加上 RDF（Resource Description Framework）处理组件。
- (2) 数据结构：应该充分考虑数据结构的可扩展性，因为 Ontology 对于扩展的不确定性，所以必须能够保证用户随时可以添加他们想要添加的内容。
- (3) 设计过程中认真考虑实验结构。
- (4) Ontology 必须经过仔细的规划后规范的被制作和维护，而不是简简单单的依赖于 Ontology 的规则。

5. 语义网络

语义网络是指一种使用网络计算的方式，在这个过程中，信息、计算资源以及服务被通过可以为计算机处理的标准的方式来描述。语义网络使资源的自动发现与组合变得更加容易，这对于资源的结合与虚拟组织的建立有很大帮助。资源的描述以及元数据通常都使用了语义网（SemanticWeb）的相关技术来实现，比如资源描述框架 RDF。

Geodise 项目采取了基于语义网络的知识管理。目标是帮助工程师在 EDSO（Engineering Design Search & Optimisation）领域进行工作，利用网络以及工程设计搜索包来更加高效的实现优化设计。为了获取需要的，能够被即使是缺乏经验的工程师也可以有效重用的，用来描述功能以及工作流的知识，Geodise 提出了分层语义网络基础架构，由以下 4 个部分组成，如图 6-18 所示。

¹ J. G. Frey etc, Less is More, Lightweight Ontologies and User Interfaces for Smart Labs.[EB/OL]. (2008-8-20)
<http://eprints.soton.ac.uk/9777/08/hughes.ppt>

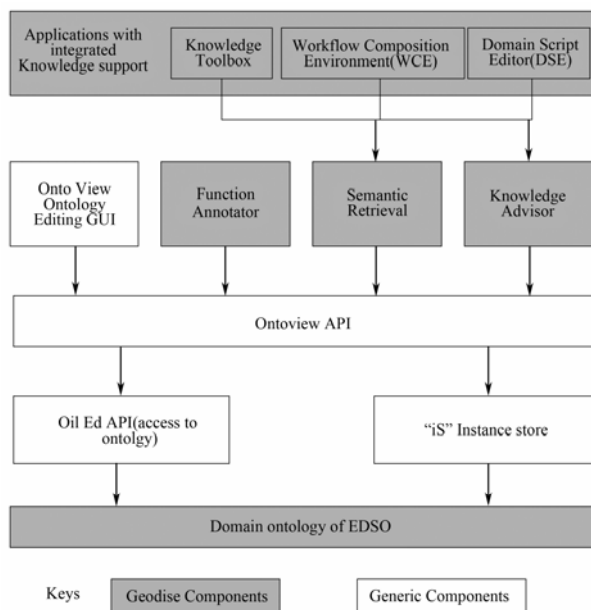


图 6-18 分层语义网络基础架构图¹

(1) EDSO 领域 Ontology

EDSO 是工程设计搜索与优化。这部分是某个针对领域而构建起来的概念及其之间联系的 Ontology，通过知识获取 (Knowledge Acquisition, KA) 的方式来获得。构建的 Ontology 需要一种标准的描述语言来描述，比如使用 DARPA 代理标识语言 DAML (The DARPA Agent Markup Language)，DAML 的目标是开发一种语言以及相应的工具来帮助界定语义网络的概念。

(2) 通用 Ontology 操作和存储机制

设计和实现通用的对 Ontology 进行操作的 OntoView 的 API 以及编辑器，在 OntoView 中进一步对 Ontology 进行优化与丰富。

(3) 基于网络的工程优化与设计搜索 Geodise 的知识服务与验证

这里包括函数标注工具、语义检索图形用户接口以及知识咨询服务。

(4) 基于网络的工程优化与设计搜索 Geodise 的应用与知识的整合

这个部分包括 Matlab 中的知识工具箱，WCE (Workflow Construction Environment) 独立工具，DSE (Domain Script Editor) 独立工具等。

在这张分层语义网络基础架构图中，部分功能说明如下：

- **OntoView:** 是一个通用知识开发与管理环境，它被用来开发 Ontology 以及支撑其他更多领域的特殊组件。OntoView API 是其提供的接口函数。

¹ Feng Tao etc, Applying the Semantic Web to Manage Knowledge on the Grid,[EB/OL].(2008-8-20), <http://www.nesc.ac.uk/events/ahm2004/presentations/189.ppt>

- 函数标注器：一种 Matlab 中的函数标注机制，用来将函数与概念进行联合。
- 语义检索 GUI（Graphical User Interface）：一种允许工程师基于 Ontology 驱动搜索标准列表来定位恰当的函数的机制。
- workflow 构造顾问：一种基于图形用户界面的组件，它和 Geodise workflow 构造环境以及一个领域脚本编辑器相整合，其中这个领域脚本编辑器使用函数的语义信息来为函数组装、 workflow 整合以及配置工作推荐语义上匹配的候选者。
- Ontology：它是由一个在 Geodise 中的 EDSO 工程专家和知识工程师协作完成的。这个标注集使用了 Ontology 在语义上描述在领域中有效的资源，比如网格可以使用的功能，优化方法以及它们的配置参数。这里的 Ontology 是通过基于 RDF 的 DAML+OIL 语言表示的。

语义网格的建造和维护过程离不开知识，知识的整个生命周期渗透在语义网格的始末。知识的生命周期包含以下四个部分（见图 6-19）。

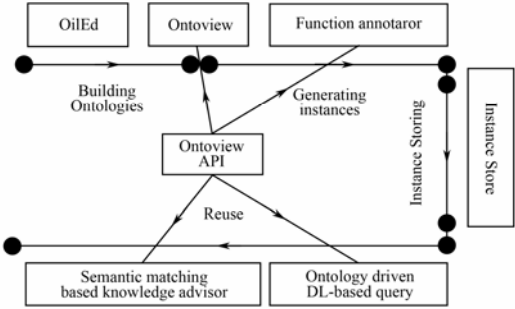


图 6-19 知识的生命周期图¹

（1）知识捕获

包括知识获取以及构造 Ontology。其中知识获取来自领域专家或者领域中的文档。而 Ontology 包含两部分，一部分是相关概念，一部分是概念之间的关系，关系要涉及层次划分，链接类型以及数据项。最终的捕获结果要用标准格式来表示，比如 DAML+OIL 格式。

（2）知识绑定

包括创建本体概念的实例——标注集。支持这个过程的工具具有 OntoView 的编辑器以及函数标注器。其中，构建语义实例的过程是通过绑定语义内容与 Ontology 实现的，其间还要丰富基于知识库的语义网络。OntoView 编辑器具有非常高性能的灵活性，可以产生任何 Ontology 概念的实例，但是使用过程有时候会令人感觉比较单调枯燥。通过一个定制过的函数标注器，可以实现自动的分析，但是这个工具缺乏灵活性，目前处理能力有限。

¹ Feng Tao etc, Applying the Semantic Web to Manage Knowledge on the Grid,[EB/OL].(2008-8-20), <http://www.nesc.ac.uk/events/ahm2004/presentations/189.ppt>

(3) 知识建模

指有用知识的详细规范：包括语义检索，函数配置与组合的建议等。

(4) 知识重用

基于语义的函数查询以及知识建议提供组件。其中函数查询需要基于它们的语义标准来识别函数，知识建议功能需要能够检索语义网络、组装函数以及配置函数。

这张知识的生命周期图（图 6-19）描述了从构建 **Ontology**、构建实例，到实例储存以及实例重用的整个知识生命周期，并且可以清楚地看到在相应的生命周期中每个阶段所涉及的主要语义网络技术。

如果想要充分收获语义网络的最大效能，比如有效的知识发现，基于计算机的互操作、自动化，网络的资源共享等，就要开发一些具体的应用实现，图 6-19 就是一个基于语义服务的具体应用情景。在图 6-20 中，通过 **Ontology** 以及其相应工具构建出 workflow 仓储，之后在网络的架构以及支持语义知识的中间件之上，构造具体的语义应用。

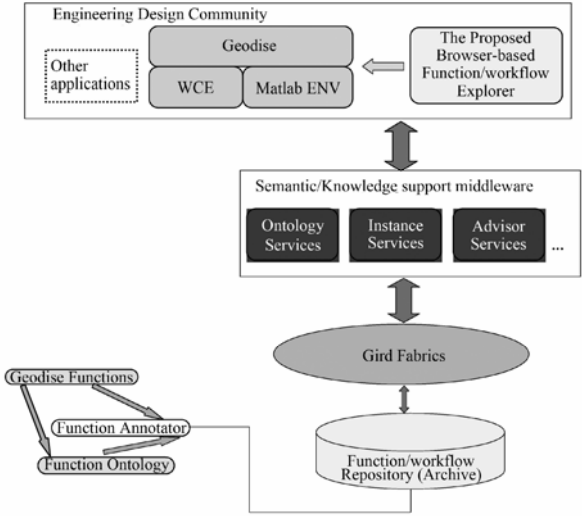


图 6-20 面向服务的语义应用情景图¹

6.3.2 网格门户技术

网格是地理位置上分布的计算机系统、数据库等各种软硬件设备通过高速互联网络融合构成的并行计算资源。网格环境中包含着各种异构计算机、软件、数据库等，要将这些资源耦合使用，使它们协同工作，需要对复杂的网格环境有清晰的认识和配置。

¹ Feng Tao etc, Applying the Semantic Web to Manage Knowledge on the Grid,[EB/OL].(2008-8-20), <http://www.nesc.ac.uk/events/ahm2004/presentations/189.ppt>

网格门户（Grid Portal）正是为了满足这种需要而产生的，它通过相对简单而且适合使用习惯的 Web 交互界面，为用户提供可视化的资源视图和快速利用网格的计算能力，对用户隐藏底层网格系统的细节，使得用户的工作不受网格变化的影响，使用网格的 Web 门户为用户提供了一个惯用的、一致的方法，使用标准的 Web 浏览器与复杂的底层网格工具和服务进行交互。

1. 网格门户和e-Science环境

Dennis Gannon提出了一个包括门户的网格计算环境结构图。他认为，一个良好的网格计算环境应当由以下几个层次组成¹。

第一层网格门户层。面对用户，具有发起、配置和控制远程网格应用实例的能力，可能需要通过一个代理服务（Factory Service）

第二层是 OGSA 层（OGSA Layer）。它包括以下通用的服务：注册和名字绑定、策略、数据管理、预留和计划、事件和消息、管理和监测、日志、服务协调、帐户管理和统计。

第三层是 OGSi 层（OGSi Layer）。它可以连接基础设施。

第四层是资源层（Resource 层）。连接远程计算机和数据服务。

Marlon Pierce 提出的基于 Grid 的门户框架。具体如图 6-21 所示¹。

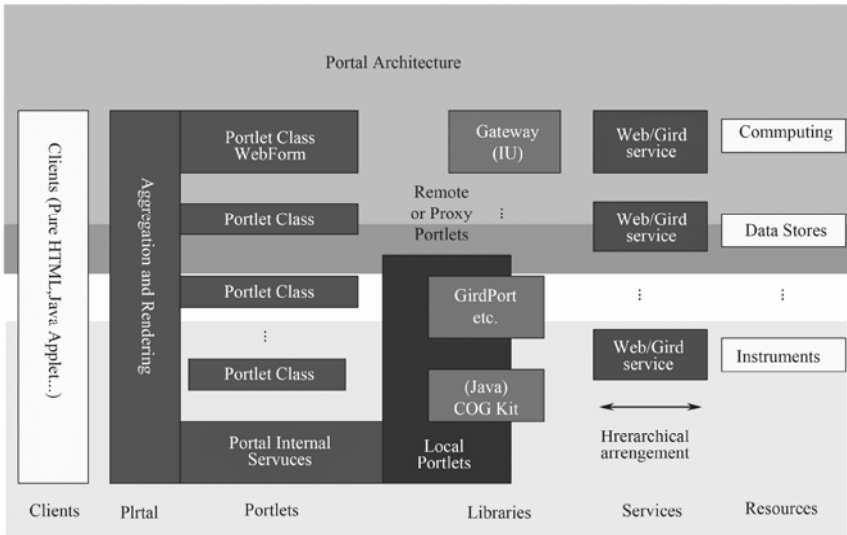


图 6-21 Marlon Pierce 提出的一个基于 Grid 的门户框架

¹ Dennis Gannon. The Computational Data Center – A Science Cloud[EB/OL].[2009-02-17].[http:// www.extreme.indiana.edu/~gannon/science-data-center.pdf](http://www.extreme.indiana.edu/~gannon/science-data-center.pdf)

Marlon Pierce 认为基于 Portlet 的门户技术是管理所有用户界面的重要方式。它可以组织许多远程或本地的“无头”服务，将这些服务集成到一个门户之中，实现针对不同用户的定制、存取控制，并且完成界面的可重用和共享。

2. 网络门户的主要功能

目前在全世界运行的网络门户有很多，其中具有代表性的有 GridPort、GILDA、GPDK、GridSphere、GRAPPA 等，它们都提供了友好的 Web 界面和增强的网格服务。它们主要有以下特点：高可靠的多用户的 Web 应用服务器；使用通用的 Web 客户技术访问；提供对远程网格资源和服务的安全访问；对各种服务提供一致的访问界面；对网格资源和服务提供全面的视图；服务于特定的网格用户团体。

网络门户一般提供以下的服务：会话管理；登陆/ 认证（authentication）；基于政策的授权（authorisation）；查询与结果；角色管理；事务管理；基于查询的资源发现；基于查询的应用选择；问题规范说明；服务注册查询服务；服务部署；生命周期管理；工作流选择及制定；资源时序安排；工作提交、执行、监控、扼要概括以及交互；事件通知服务；视频/音频传送以及基于门户的协作服务；其中核心服务如图 6-22 所示。

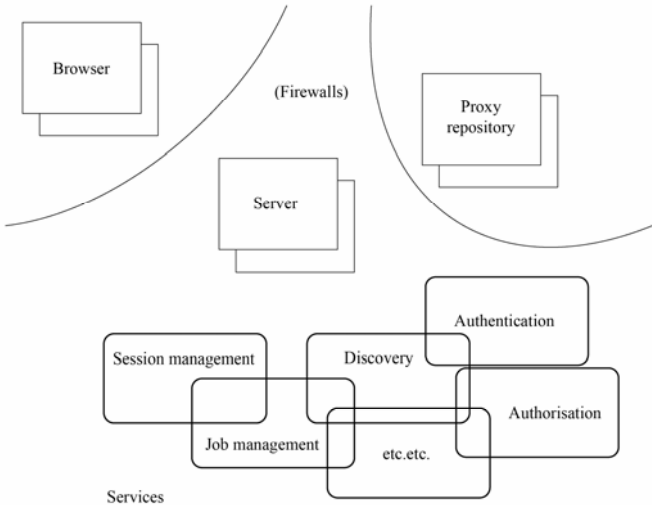


图 6-22 网络门户核心服务¹

支持 Grid Portal 的主要技术标准有 JSR-168 和 WSRP。其中 JSR-168 Java portlet 规范、WSRP 是基于 Web Service 的 portlet 集成规范。二者关系如图 6-23 所示。

¹ Rob Allan, Chris Awre, Mark Baker & Adrian Fish, Portals and Portlets 2003, published Sep 2004, [EB/OL].(2008-8-20), http://www.nesc.ac.uk/technical_papers/UKeS-2004-06.pdf

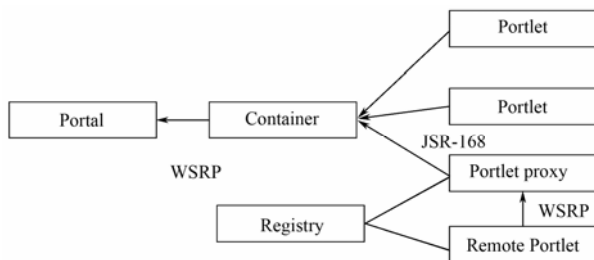


图 6-23 JSR-168 和 WSRP¹

3. JSR 168: Java Portlet规范

现在大部分门户软件都支持 Portlet 技术。为了解决多种因为不兼容的接口产生的应用程序提供者、Portal 消费者和 Portal 服务器厂商之间的问题，在 IBM 和 Sun 共同推动下，Java Community Process (JCP) 于 2004 年 10 月发布了 Java Portlet 规范——JSR (Java Specification Requests) 168，从而提供了 Portlet 和 Portal 之间的接口标准。本质上看，JSR 168 解决的是本地 portlet 与门户的接口问题。一个 portlet 只要是符合 JSR 168 标准，无论是谁开发的，都能很轻松地支持 JSR 168 的门户采用，无须经过特别的编程。

4. WSRP (WebServices for Remote Portlets, 基于Web服务的portlet集成规范)

将远程内容和应用集成到门户中也是一项要求进行定制编程的繁重的任务。不同的应用和内容提供者采用了各不相同的接口和协议，为了使门户能够与这些提供者进行通讯，运行门户的提供商或组织者不得不编写多个单独的适配 portlet 才能集成远程的内容和应用。WSRP 标准简化了远程应用/内容集成到门户中的过程，使门户管理员可以轻松地从丰富的远程服务中选择所需的服务并集成到自己的门户之中，而无须进行编程。

WSRP 引进了 portlet 生产者和 portlet 消费者的概念：消费者聚合 WSRP 兼容的 portlet 到一个门户中，并管理它与终端用户的交互；生产者驻留并管理基于 WSRP 标准的供消费者调用的 portlet (如图 6-24 所示)。生产者的模型是 portlet 容器。生产者的服务接口包括：自描述 (Service description)、标识片断、注册和 portlet 管理。自描述接口用于提供关于生产者所提供的服务相关的信息，是必备接口。标记片断 (Markup) 接口也是一个必备接口，用于请求/交互标识片断。注册 (Registration) 接口则是一个可选的用于提供消费者在生产者处的登记管理的接口。portlet 管理 (Portlet management) 接口用于控制 portlet 的复制和定制，同样是可选接口。

¹ Rob Allan, Chris Awre, Mark Baker & Adrian Fish, Portals and Portlets 2003, published Sep 2004,[EB/OL].(2008-8-20), http://www.nesc.ac.uk/technical_papers/UKeS-2004-06.pdf



图 6-24 portlet 生产者和 portlet 消费者之间的关系

生产者可以有选择地管理消费者的注册信息，并且可以要求他们在调用 portlet 之前必须进行注册，通过注册确定消费者和生产者之间的合作关系。消费者本质上近似于以终端用户名义实施的路由器，消费者将用户的请求指向合适的生产者，接着一个生产者处理该请求并将结果返回给消费者，消费者聚合来自多个生产者的返回结果并将聚合结果返回给用户。既然消费者起着终端用户代理的作用，那么许多用户的特定信息将通过消费者传递。消费者必须区分不同用户的通信信息，并且必须保证交互过程中特定用户的隐私性。

图 6-25 描述了支持 WSRP 的门户之间的交互过程：①消费者发现一个生产者；②消费者和生产者建立关系；③消费者学习生产者的所有功能；④终端用户和消费者建立联系；⑤消费者为用户聚合多个页面，经常通过 portlet 实现；⑥终端用户向消费者提供页面请求；⑦消费者向生产者请求信息；⑧生产者为消费者提供逻辑和数据；⑨终端用户看见聚合后的页面。

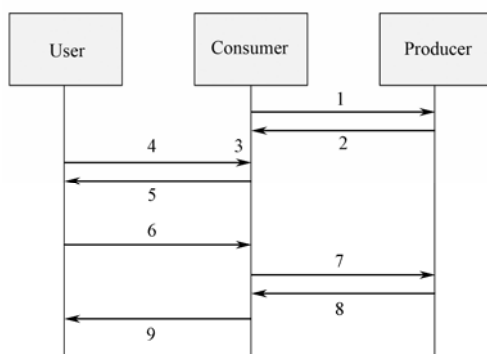


图 6-25 支持 WSRP 的门户之间的交互过程

用户、生产者和消费者三者之间 WSRP 服务的实施过程如图 6-26 所示。

WSRP4J 项目提出了 WSRP4J 生产者（WSRP4J Producer），在由 Tomcat，Axis 和 WSRP4J 构成的一组开放源软件环境中提供 WSRP 兼容的服务。WSRP4J 已实现了与 JSR 168 参考实现——Pluto 的融合。另外，如图 6-27 WSRP4J 还提供了 WSRP4J 消费者（WSRP4J Consumer）作为写到 Portlet API 的通用代理 portlet。

5. 支持Portlet的开放源门户软件

一个普通的网站系统要支持 portlet，其做法可以是在已有 portlet 容器的基础上自行开发门户系统，但更简便的方法是直接采用已有的功能齐全的支持 portlet 及其相关标准的门户软件。本节将对 portlet、对 JSR168 标准和 OASIS WSRP 标准的支

持程度以及该软件的实用性和易用性几个方面来比较部分现有的一些门户软件。目前已有多种开放源门户软件，如 Pluto、Jetspeed、liferay 和 uPortal。

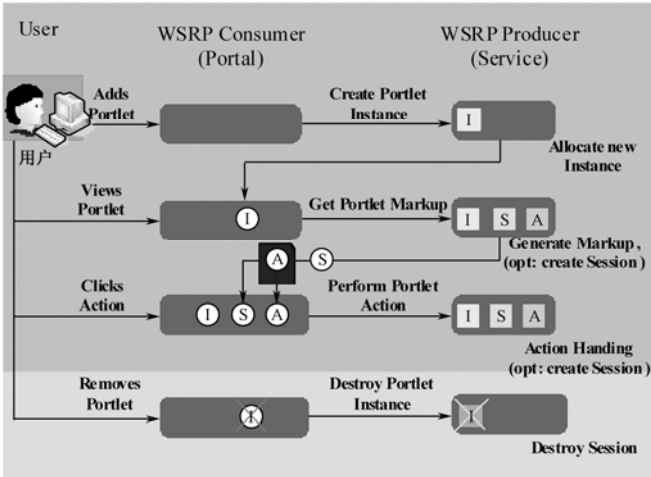


图 6-26 用户、生产者和消费者三者之间 WSRP 服务的实施过程

6. GridPort工具包

GridPort 汇集了网格软件包中的网格服务，并对外提供其他网格功能，同时还为门户和应用程序开发人员提供简单、一致的 API。GridPort 集成了其他网格软件中提供的网格服务，例如 Globus 和网络天气预报（Network Weather Service），而且还提供了其他的网格服务，以便更简单有效地利用网格资源。GridPort 消除了开发者学习每种底层网格技术和服务的 API 的需要。图 6-28 展示了充当界面与底层网格应用程序间中间件的 GridPort。

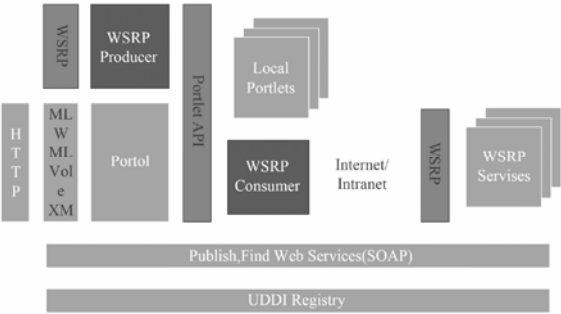


图 6-27 WSRP4J¹

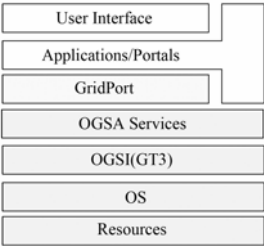


图 6-28 GridPort²

¹ WSRP Architecture,[EB/OL].(2008-8-20),<http://portals.apache.org/wsrp4j/arch/index.html>
² Tomislav Urban,GridPort: Using Globus for Grid-Enabled Web Portals, [EB/OL]. (2008-8-20), http://gridport.net/main/presentations/GridPort_3_GlobusWORLD_04.ppt

GridPort 为门户和应用程序提供了信息和交互式服务（图 6-29）。它使用信息服务提供与网格有关的数据，这些数据从底层资源的数量、状态以及处理能力有关的各种网格技术中获得。**GridPort** 交互式服务可以为产品提供很多功能，如文件管理、命令执行和任务提交。**GridPort** 的关键组件有：

（1）**GridPort 信息存储库（GridPort Information Repository, GPIR）**：它搜集并记录“动态”和“静态”的网格数据，并通过 Web 服务为门户和应用程序提供当前数据和历史数据。

（2）**身份验证**：**GridPort** 使用 **Globus Grid Security Infrastructure (GGSI)** 对远程主机进行单一的签名和身份验证。

（3）**文件管理**：**GridPort** 的文件管理功能允许在底层的网格系统上执行与文件有关的任务。

（4）**作业管理**：**GridPort** 为作业管理提供了几种功能。其中有两种执行作业的形式：批处理形式和交互形式。

（5）**Job Sequencer**：**Job Sequencer** 支持任务按照线性顺序执行，否则它们只能单独执行任务，**Job Sequencer** 允许执行预先定义的任务流。

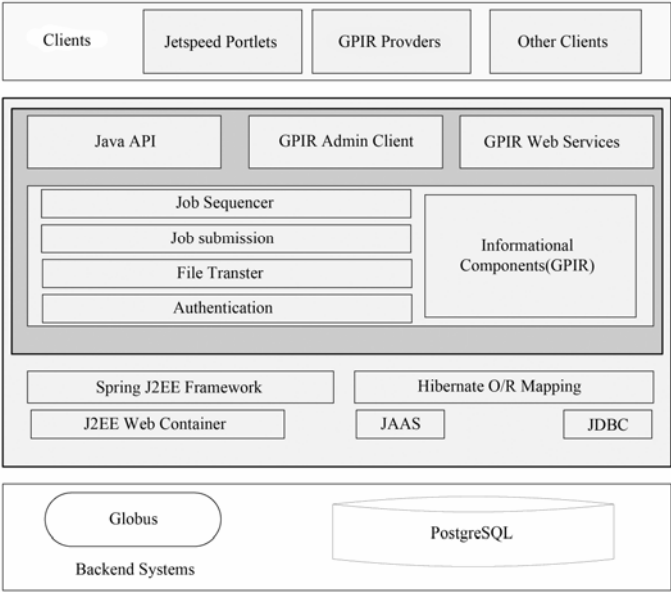


图 6-29 GridPort架构¹

¹ Maytal Dahan, Eric Roberts, GridPort 3 Training,[EB/OL], (2008-8-20), <http://gridport.net/main/raining/10.25.04.GridPort3Training.ppt>

6.4 Globus

Globus ToolKit 工具包来源于 Globus 项目。Globus 项目是国际上最有影响力的与网格计算相关的项目之一，是由来自世界各地关注网格技术的研究人员和开发人员共同努力的成果。Globus Toolkit 是一个开放源码的网格基础平台，基于开放结构、开放服务资源和软件库，并支持网格和网格应用，目的是构建网格应用提供中间件服务和程序库。

Globus Toolkit 具有较为统一的国际标准，有利于整合现有资源，目前大多数网格项目都采用基于 Globus Toolkit 所提供的协议及服务建设。

自从 1997 年起，Globus ToolKit 工具包的第二版（GT2）成为了网络计算的事实标准。它着重于可用性和互操作性能力，定义和实现了一些协议、API 和服务。GT2 通过提供授权认证、资源发现和资源访问等共同问题的解决方案，加快了网格应用的构建。但 GT2 毕竟针对网格的具体实现，所以它并没有一个正式的标准。

2002 年 2 月，在加拿大多伦多市召开的全球网格论坛 GGF 会议上，Globus 项目组和 IBM 共同倡议了一个全新的网格标准 OGSA。OGSA 把 Globus 标准与以商用为主的 Web Services 的标准结合起来，网格服务统一以 Services 的方式对外界提供。2003 年符合 OGSA 规范的 Globus Toolkit 3.0（GT3）发布，这标志着 OGSA 已经从一种理念、一种体系结构进入实践阶段。GT3 提供一个完整的开放网格服务基础设施实现，它的许多功能重构成与 OSGI 兼容的服务。

2004 年 1 月发布了统一网络计算和 Web 服务的新标准“WS-Notification”和“WS-Resource Framework”。Web 服务资源框架是 OSGI 的重构和发展，利用新的 Web 服务标准。

WSRF 基本保留了 OSGI 中的所有功能，同时更改了一些语法，并且还在其表示中采用了不同的技术。Web 服务通知为 Web 服务提供基于消息发布和预定能力。WSRF 和 WSN 都是建立在已存在的 web 服务定义和技术基础上的，帮助实现了网格计算、系统管理和 Web 服务的统一。

Globus ToolKit 4（GT4）实现了 WSRF 和 WSN 标准。GT4 提供 API 以构建有状态的 Web 服务。其目标是建立分布式异构计算环境。所有知名的 GT3 协议通过重新设计均可使用 WSRF，此外，GT4 也在其中增添了一些新的 Web 服务的组件。

6.4.1 GT3（Globus Toolkit3）

1. GT3 概述

2003 年 6 月发布的 GT3 是第一个可以支持 Web 服务的 Globus Toolkit 版本，它利用了 OSGI，这是 WSRF（WS-Resource Framework）的前身。不幸的是，OSGI 不

是一个业界普遍认可的标准，因此，坦白地说，GT3 软件并没有达到产品级别的质量。GT3 的体系结构见图 6-30，图中白色的部分是 GT3 Core 提供的。它们是建立网格服务的基础。OGSI 参考实现（OGSI Reference Implement, RI）完成了 OGSI 规范 1.0 所定义接口的实现，以 API 和工具的形式提供给用户以方便开发与 OGSI 兼容的网格服务。安全基础设施（Security Infrastructure, SI）提供 SOAP 安全、传输层安全、相互认证、单点登录和服务认证等功能。GT3 核心还提供了系统级服务作为运行时和其他服务相关联的基础，它们建立在 OGSI-RI 和 OGSI-SI 之上，GT3 同时还提供了一些基础服务，如程序执行，数据管理和信息服务等。这些服务是建立在 OGSI 和 GSI 组件之上的。用户定义服务是指由用户开放的高层服务，它不由 GT3 提供，但可以建立在任何 GT3 的组件之上，包括基础服务。所有这些服务是和抽象的 OGSI 运行时环境相交互的，称为网格服务容器。网格服务容器把应用和服务运行时细节分离开来，同时还控制服务的生命期，把消息分发给对应的服务实例等。容器的前端封装了一个标准的 Web 服务引擎接口，用来实现 XML 消息映射。最后，Web 服务引擎和网格服务容器宿主到一个主机环境中。

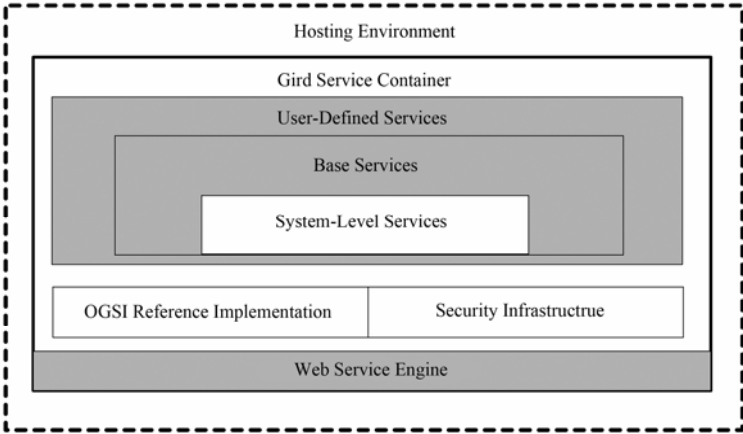


图 6-30 GT3 的体系结构图¹

2. GT3 支持的主机环境

目前 GT3 支持 4 种 Java 主机环境：

- (1) **Embedded**：以库文件的形式将 OGSI 主机环境嵌入已有的 J2SE 应用中。
- (2) **Standalone**：一个可以部署网格服务的轻量级的 J2SE 服务器。
- (3) **J2EEWebContainer**：将 OGSI 主机环境嵌入 Web 服务器中，该 Web 服务器具有 Java Servlet 引擎，如 Jakarta Tomcat。

¹ GT3 core architecture, [EB/OL].(2008-8-20), <http://www.globus.org/toolkit/docs/3.2/core/key/index.html>

(4) J2EE EJB Container: 把网格服务容器嵌入到 EJB 应用服务中。

3. GT3 服务器端的组件

图 6-31 说明了 GT3 服务器端的组件，服务器端框架的主要体系结构组件包括：

- (1) Web 服务引擎。该引擎由 Apache AXIS 框架软件提供，用于处理一般的 Web 服务行为、SOAP 消息处理、JAX-RPC 处理程序处理和 Web 服务配置。
- (2) Globus 容器框架。GT3 提供了一个容器，它通过唯一的实例句柄、实例资源库和生命周期管理（包括服务激活/钝化和软状态管理）管理有状态的 Web 服务。

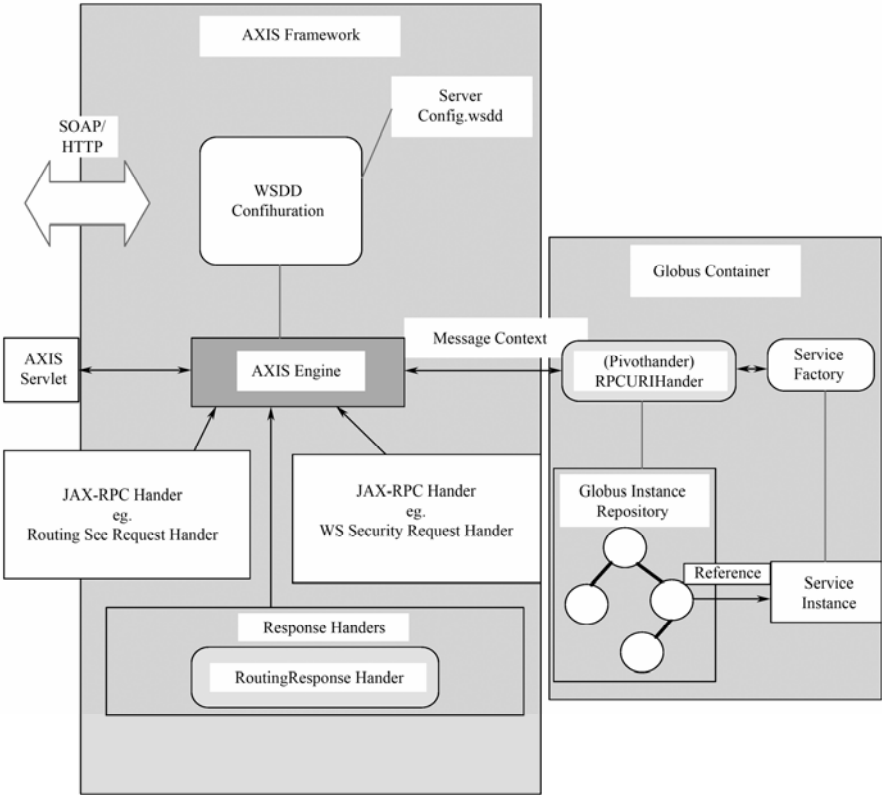


图 6-31 GT3 软件框架：服务器端体系结构组件¹

目前，GT3 将 Apache AXIS 用作其 Web 服务引擎，Apache AXIS 在 J2EEWeb 容器中运行并提供 SOAP 消息侦听器（AXIS servlet）。它负责 SOAP 请求/响应序列化和反序列化、JAX-RPC 处理程序调用和网格服务配置。如图 6-31 所示，GT3 容器

¹ Joshy Joseph, Globus Toolkit 3.0 和 OGSi 体系结构-概述,[EB/OL].(2008-8-20), <http://www.ibm.com/developerworks/cn/grid/gr-gt3/>

向 **AXIS** 框架提供枢轴（**pivot**）处理程序来将请求消息传递给 **Globus** 容器。

这个容器体系结构用于管理 **Web** 服务的状态性质以及这些服务的生命周期（网络服务实例是有状态的 **Web** 服务）。一旦服务 **foctory** 创建了网络服务实例，框架就为该实例创建唯一的网络服务句柄（**GSH**），然后向容器资源库注册该实例。该资源库拥有所有有状态的服务实例，由其他框架组件和处理程序与之连接来执行以下服务：

- （1）标识服务和调用方法。
- （2）获取/设置服务特性（如实例 **GSH** 和 **GSR**）。
- （3）激活/钝化服务。
- （4）解析网络服务句柄以引用服务并持久存储它。

图 6-32 说明了客户机端的组件。如 图所示，**Globus** 使用标准的**JAX-RPC**客户机端编程模型和**axis**客户机端框架网络服务客户机。除了标准的**JAX-RPC**编程模型外，**Globus** 还在客户机端提供许多助手类，以隐藏**OGSI**客户机端编程模型的细节。

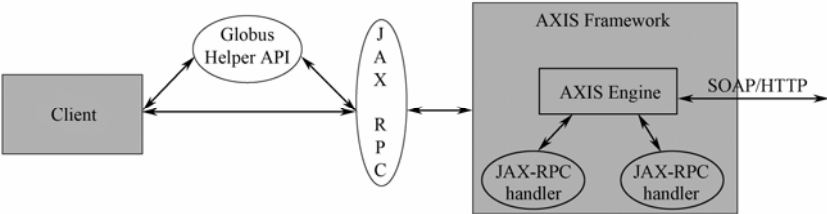


图 6-32 GT3 客户端组件图¹

6.4.2 GT4（Globus Toolkit4）

从GT3 发布以后的两年的时间里，**Globus**已经在很多前沿取得了进步。一些重要规范也已经在GGF、OASIS和W3C中得到了改进——特别是 **WSRF**。2005 年**Globus Alliance**发布了GT4。

GT4 是一组支持创建 **SOA**（**Service Oriented Architectures**）的核心功能工具。这些核心功能包括配置计算和数据资源的执行和数据管理功能——大量数据的移动和管理，以及资源发现和监视的服务的支持。

图 6-33 是 GT4 中所包含的组件模块。

¹ Joshy Joseph,Globus Toolkit 3.0 和 OGSI 体系结构-概述,[EB/OL].(2008-8-20), <http://www.ibm.com/developerworks/cn/grid/gr-gt3/>

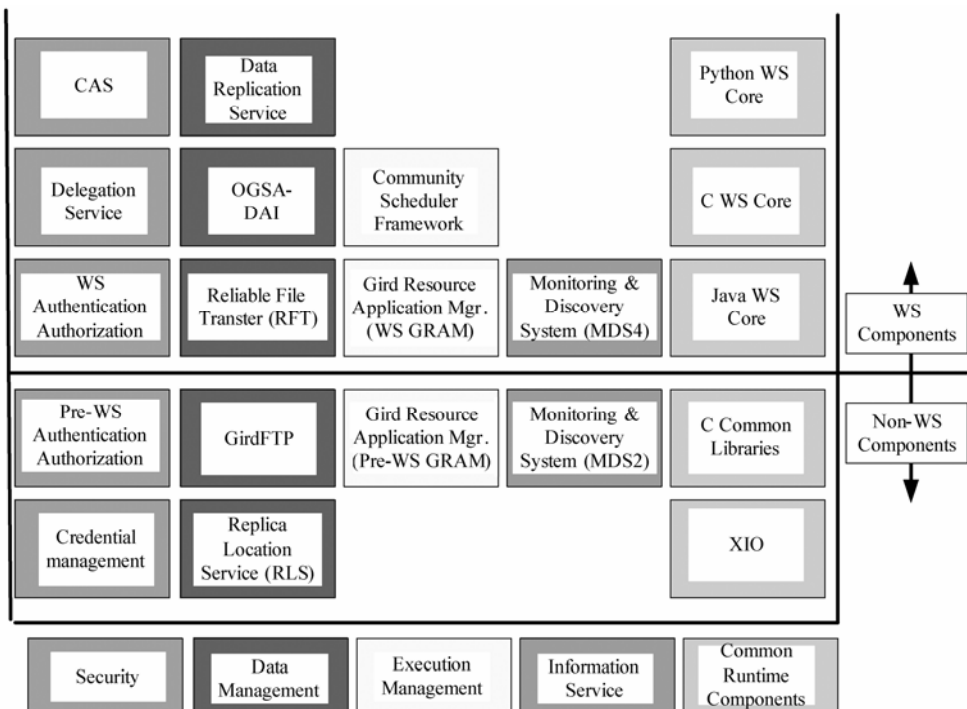


图 6-33 GT4 中所包含的组件模块图¹

GT4 是 Globus 的一个延续。GT4 中已经实现了最新的 Web 服务标准，包括 WS-Security 和 WSRF。GT4 现在已经达到了产品级别的质量——代码质量足以满足企业需求。

1. Globus Toolkit 4.0 的新特性

GT4 中最引人注目的特性是：与 GT 以前的版本相比，GT4 包括很多 Web 服务和标准。GT4 基于 Web 服务实现了对各种资源类的标准接口，这一事实可以显著简化设计，从而增加这些标准被接受的可能。当然，通过开源社区开发人员和 Globus 社区中的用户的努力，与之前的版本相比，GT4 在测试质量、文档、性能和可靠性方面都得到了很大提高。除此之外，GT4 提供了以下一些特性：

- GT4 遵守最新的 WebServices Interoperability Organization (WS-I) Web 服务标准，在不同环境之间提供了最大的互操作性。
- GT4 包含对一些重要授权标准的初步支持，其中包括 Security Markup Language (SAML) 和 eXtensible Access Control Markup Language (XACML)。

¹ Components in the Globus Toolkit,[EB/OL].(2008-8-20), <http://www.globus.org/toolkit/docs/4.0/GT4figure.jpg>

- GT4 实现了 WSRF 和 WS-Notification 规范。
- GT4 具有复杂的授权和安全功能。Globus 在网络安全性方面一直非常努力，从安全性的观点来看，GT4 已经是满足企业需求。
- GT4 提供了高级的执行和数据管理功能。

2. GT4 中服务器端的主要组件结构

图 6-34 主要描述了在服务器端的GT4中的主要组件结构，这只是GT4 所提供功能的一部分。GT4 结构由一个网络容器组成，网络容器主要用来管理所有部署的WEB 服务，贯穿于每个WEB服务的运行周期。GT4 使用 apache的axis作为它的WEB服务的引擎来处理所有的SOAP消息，JAX-RPC（Java API for XML-Based RPC）处理和WEB服务的配置。

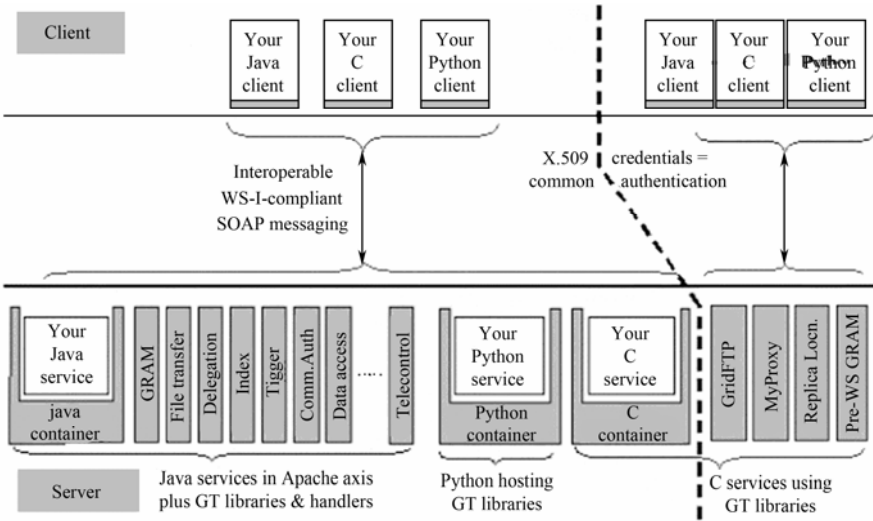


图 6-34 GT4 中服务器端的主要组件结构图¹

3. GT4 的安全特性

在任何网络环境中，安全性都是一个非常重要的问题。这种需要对网络环境来说尤其重要，因为在网络环境中，客户机可以位于不同的地理空间和组织中，而 GT4 可以满足这种需求。任何网络对于安全性都有一些基本的要求，这些要求包括但不限于以下几点：

- 保证网络实体之间的通信是安全的、可以防止篡改的，这些网络实体包括用

¹ I. Foster. Globus Toolkit Version 4: Software for Service-Oriented Systems. IFIP International Conference on Network and Parallel Computing, Springer-Verlag LNCS 3779, pp 2-13, 2006.

户、资源和程序。

- 网格用户在多个资源之间使用单点登录的能力。
- 从一个实体到另外一个实体的权限委托，用于类似代理的操作。

目前 GT4 为安全性提供的是一组基于标准的授权和身份验证功能的开源实现。GT4 实现了 WS-Security 及其相关规范，这使得用户在应用程序中只需进行一次身份验证，就可以访问整个企业或其他企业中的计算资源和数据资源，而不需要再进行任何身份验证。但 GT4 目前尚未提供的功能是整合企业的单点登录解决方案。

GT4 中的安全组件是 GSI (Grid Security Infrastructure)，GSI 允许网格实体在开放网络上进行身份验证、授权和安全通信。GSI 是 GT4 的一个组件，用于解决所有这些需求，为网格通信提供保密性、完整性和回放保护（为了防止监听和中间人攻击），并为网格用户提供单点登录和权限委托。此外，它还包括一些用来对网格实体的身份进行验证的工具，并基于这些工具确定允许该实体执行哪些操作。

在 GT4 中，GSI 可以同时支持传输层的安全性和消息级的安全性。在 GSI 中，传输层的安全实现可以在网格实体之间使用 TLS (Transport Layer Security) 和 X509 证书进行身份验证。消息级的安全性实现了对 WS-Security 和 WS-SecureConversation 规范的支持。通过使用这种安全性，可以在每个消息的级别上为 SOAP 消息提供安全保护。

4. GT4 的应用

GT4 已经应用在一些项目中了，比如：

- 南加利福尼亚大学信息科学研究所等机构主持的 Planning the SCEC Pathways: Pegasus at work on the Grid 项目。
- 科罗拉多大学 Boulder 分校等机构主持的 Grid-BGC: A Grid-Enabled Terrestrial Carbon Cycle Modeling System 项目。
- USC 信息科学研究所等机构主持的 Experience in Building the NEESgrid: Lessons Learned 项目。
- 南安普顿大学等机构主持的 Grid State of the Users: 25 Conversations with UK eScience Projects 项目。

5. GT4 组件研究

GT4 中包含的组件一共分为 5 类，分别是安全、数据管理、运行管理、信息服务以及常见运行组件。这 5 类组件每个类别下又包含具体的组件服务，如表 6-2 所示。

表 6-2 GT4 包含组件

类别	包含组件
安全	1. Community Authorization Service (CAS) 2. Delegation Service 3. WS Authentication Authorization 4. Pre-WS Authentication Authorization 5. Credential Management
数据管理	1. Data Replication Service 2. OGSA-DAI 3. Reliable File Transfer (RFT) 4. GridFTP 5. Replica Location Service (RLS)
运行管理	1. Community Scheduler Framework 2. Grid Resource Allocation Mgr (GRAM) 3. Grid Resource Allocation Mgr (Pro-WS GRAM)
信息服务	1. Monitoring & Discovery System (MDS4) 2. Monitoring & Discovery System (MDS2)
常见运行组件	1. Python WS Core 2. C WS Core 3. Java WS Core 4. C Common Libraries 5. XIO

(1) 安全

① CAS (Community Authorization Service) 社区认证服务

CAS 允许虚拟组织对分布在不同地点的资源的制定和发布策略。一台 CAS 服务器向虚拟组织用户陈述声明与规则，从而保证它们能够得到恰当的访问资源的权限。除此之外，CAS 服务器还识别并加强这些声明与规则。目前 CAS 已经测试过的运行平台包括：Windows XP、Linux (Red Hat 7.3)。测试过的容器包括：Java WS Core container、Tomcat 5.0.30。CAS 能够扩展适用于多种服务，目前被 GridFTP 服务所支持。

CAS 服务依赖于以下的一些 GT 组件或第三方软件组件：WS Authentication and Authorization, Java WS Core, Pre-WS Authentication and Authorization, OpenSAML, Libxml。在 GT4.0 中 CAS 没有提供新特性，CAS 支持的其他特性包括对 GridFTP 的文件级访问控制和 SAML (Security Assertion Markup Language) 认证决定的发布。

② Delegation Service

Delegation Service 是 Globus Toolkit 4.0 中的一个新组件。这个组件为群集环境

提供了一个可信任的代理接口。在群集环境下，它使得一个单一的可信任代理可以被多个服务请求者使用，例如，它可以被多个 GRAM 工作请求或者跨 GRAM 和 RFT（Reliable File Transfer Service，可靠文件传输服务）请求所使用。它同时也为保密更新提供了一种方式。目前 Delegation Service 已经测试过的运行平台包括：Windows XP、Linux（Red Hat 7.3），测试过的容器包括：Java WS Core container、Tomcat 5.0.30。

在 GT4.0 中，Delegation Service 提供了二个主要功能特性：第一个是为主机提供了一个代理接口和保密更新，第二个是允许多个服务请求者重用单一可信任代理。Delegation Service 主要依赖于以下的 GT 组件或第三方软件组件：WS Authentication and Authorization, Java WS Core, Apache Axis。

③ WS Authentication Authorization

在 GT4 中，WS Authentication Authorization 取代了原来的 Grid Security Infrastructure（GSI）。WS Authentication Authorization 主要被分为两部分：消息级别安全（Message level security）和认证框架（Authorization framework）。

消息级别安全实现了两个标准：WS-Security 和 WS-SecureConversation。这两个标准提供了 SOAP 消息的加密、完整性验证和重放保护。WS Authentication and Authorization 传输级安全通过使用基于 SSL/TLS（HTTPS）的 HTTP 协议提供了一种消息传递的安全渠道。这种安全机制支持由 SSL/TLS 个 X.509 代理认证提供的所有安全特性。在 GT4.0 中，消息级别安全提供了以下几个新功能特性：符合已经发布的 IBM/Microsoft 的 WS-Trust 和 WS-SecureConversation 详细规范，符合 WebServices Security 1.0 标准，支持 HTTPS。消息级别安全支持的其他功能特性：消息加密、完整性保护和防止重放攻击，轻量级消息保护的 session key 建立。GT4 中被去掉的功能特性是 GT 3.2 SecureConversation 协议

认证框架组件提供了一个容器级别认证的框架。认证框架允许一系列有良好定义接口的认证模块和各种实体联合，比如容器中的各种服务。同时，认证框架还提供了多种不同认证模块实现，从基于网格地图的认证到使用 SAML 协议来查询外部服务进行认证决定的模块。目前认证框架已经测试过的运行平台包括：Windows 2000、Linux（Red Hat 7.3）和 Solaris 9。在 GT4.0 中，认证框架提供了以下新功能特性：一个 SAML callout 认证模块，该模块使外包认证决定成为一个授权服务。比如，PERMIS（Performance Metrics for Intelligent Systems）认证框架支持的其他功能特性：基于网格地图文件和其他访问控制列表的认证，实现定制认证模块的能力。认证框架主要依赖于以下的 GT 组件或第三方软件组件：WS Authentication and Authorization Message-Level Security, OpenSAML。

④ Pre-WS Authentication Authorization

Globus Toolkit Pre-WebServices Authentication and Authorization 组件提供了 API 和工具来实现认证、授权和证书管理功能。认证 API 基于 Public Key Infrastructure

(PKI) 技术, 比如X.509 Certificates and TLS。除认证之外, 还具有基于X.509 代理证书的代理机制。

授权功能支持采用了一对 API。第一个提供了一个通用的认证 API, 通过这个 API 允许插图的编号 (callouts) 来执行基于用户的信任证书的访问控制, 比如 X.509 certificate chain。第二个 API 提供了一个简单的访问控制列表, 这个列表将授权过的远程实体和本地系统用户名映射起来。第二种机制同时还提供了 callouts 授权第三方重载缺省行为, 这种机制目前被应用在 Gatekeeper 和 GridFTP 服务器中。

除了上面提到的之外, 还有各种低层次的 API 和工具用来管理、发现和查询证书。

认证框架支持的其他功能特性: 用户使用标准 X.509 终端实体和代理认证的授权, 使用 X.509 代理认证的代理, 基于 GridFTP 和 Pre-WS GRAM 的用户证书链可插入式认证, 基于工作的 RSL (Resource Specification Language) 的 Pre-WS GRAM 的可插入式认证。

⑤ Credential Management (SimpleCA & Myproxies)

Credential Management 又分为 SimpleCA 和 Myproxies 两部分。

SimpleCA 提供简单证书授权包, 目的是为了向 Globus Toolkit 的用户和服务发布信任证书。SimpleCA 支持的其他功能特性: 为 Globus Toolkit 使用的 X.509 certificates 的简单创造, 为已经创造的 SimpleCA 使用的 GPT 包简单创造。SimpleCA 主要依赖于以下的 GT 组件或第三方软件组件: Pre-WS Authentication and Authorization, OpenSSL。

MyProxy是GT4 中新增加的组件, 它是一个在线的信任仓储。可以将X509 代理信任状存储在MyProxy仓储中, 信任状由一个传输阶段保护, 并且可以在网络上检索。这就结束了在机器间手工拷贝私人密码和认证状文件的需求。MyProxy还可以被用于网格门户和项目经理的信任更新的认证。MyProxy已经测试过的平台包括: Mac OS X 10.3, i686 GNU/Linux和ia64 GNU/Linux。MyProxy支持的其他功能特性: 通过使用 myproxy-init和myproxy-logon, 用户可以储存和检索X.509 代理信任状; 通过使用 myproxy-store和myproxy-retrieve, 用户可以储存和检索多个X.509 终端实体信任状; 通过使用myproxy-admin-load-credential, 管理员可以代表用户用X.509 终端实体信任状加载仓储, 管理者可以使用myproxy-admin-adduser命令创建用户信任状将其加载到MyProxy仓储中, 用户和管理者可以对仓储中的信任状设置访问控制策略, 如果策略允许, 工作经理可以在信任状过期之前更新它们, MyProxy服务器通过使用可配置的外部call-out加强本地传输阶段策略。MyProxy主要依赖于以下的GT组件或第三方软件组件: Pre-WS Authentication and Authorization。

(2) 数据管理

① 数据复制服务 (Data Replication Service, DRS)

数据复制服务对网格文件提供了基于拉的复制能力。它是一种高层次的数据管理服务, 它建立在两种 GT 数据管理组件基础之上: RFT 和 RLS (Replica Location

Service，复本定位服务）。

数据复制服务的功能是为了保证一个文件的详细集合被存储在存储器上。DRS 通过查询 RLS 来发现被检索文件在网格上的位置。在文件被定位之后，DRS 创建一个传输请求，之后这个请求被 RFT 所执行。在传输结束后，DRS 用 RLS 注册这个新的复制品。

DRS 被实现为一个 Web 服务，并且通过 WSRF 详细规范来编译。当一个 DRS 请求被收到之后，它会创建一个用于维持每个被复制文件的状态的 WS-Resource，其中包括对这些文件的成功或者失败的操作。

② RFT

总体来讲，globus-url-copy 和 GridFTP 都是非常强大的工具，但是它们也都存在着一些问题。首先，GridFTP 协议并不是 Web Service 协议，其中没有涵盖 SOAP 和 WSDL。其次，GridFTP 要求在整个传输过程中，客户端向服务器保持一个开放的 socket 连接。对于长时间的传输，这个可能会很不方便，比如在笔记本上使用。因此还提供了一种被称谓 RFT 的服务。

RFT 是基于 Web Services Resource Framework (WSRF) 的 Web 服务，它提供了“工作安排器”，例如数据移动的功能。客户只要提供来源与目的地的 URL 列表（包含目录或文件地址），接下来服务会把客户的工作描述写入到数据库中，然后来传输文件。一旦这项服务接受了工作请求，整个交互过程和其他的工作安排器很类似。同时还提供了查询传输状态服务方法，当然也可以使用标准的 WSRF 工具（在 Globus Toolkit 中提供）来预定状态变更通知。GT4 提供了安装在 Web 服务容器或者简单客户端的服务实现，同时还为用户开发提供了 Java API。

③ GridFTP

GridFTP 是一个由 Global Grid Forum Recommendation GFD.020, RFC 959, RFC 2228, RFC 2389 定义的协议，同时也是 IETF FTP 工作组之前的一个草稿。GridFTP 协议提供了安全、健壮、快速和高效的数据传输。Globus Toolkit 提供了最常用的 GridFTP 协议的实现。

Globus Toolkit 提供了一个称为 globus-gridftp-server 的服务器实现，一个称为 globus-url-copy 的脚本命令行客户端和客户开发使用的类库集合。虽然 Globus Toolkit 没有提供可以交互的客户端程序，但是 GridFTP 用户指南却提供了至少一个由其他项目开发的交互性客户端程序。

如果用户希望为其他人提供数据，首先需要在主机上安装服务器程序，这样可以访问数据并且确保和存储数据的存储系统之间存在一个恰当的数据存储接口 Data Storage Interface (DSI)。这通常意味着一个标准的 POSIX 文件系统，但是数据存储接口也是可以的，比如存储资源代理 (Storage Resource Broker,SRB)，高性能存储系统 (the High Performance Storage System, HPSS)，以及来自威斯康星麦狄逊大学 Condor team 的 NeST。

当安装好上面的程序后，用户还需要一个 **GridFTP** 客户端来访问数据。**Globus Toolkit** 提供了一个叫做 **globus-url-copy** 客户端软件。这个客户端软件可以通过一系列的协议（比如 **http**, **https**, **ftp**, **gsiftp**）来访问数据。需要注意的是，这个客户端没有图形界面，仅仅是靠命令行来实现的，比较适合编写脚本。比如，以下的命令：

```
globus-url-copy gsiftp: //remote.host.edu/path/to/file file: ///path/on/local/host
```

这条命令将会从一个远程用户传输一个文件给本地可以访问的指定的 URL。

最后，如果客户希望访问存储在 **GridFTP** 服务器之后的文件，或者希望开发自己的客户功能，可以使用提供的强大的客户端类库来开发。

④ RLS（复本定位服务，Replica Location Service）

RLS 是一个用来追踪复制数据存放在物理存储系统位置的简单注册。

RLS 的用户或者服务在文件被创建的时候注册文件。之后，用户通过查询 **RLS** 服务器来得到这些复制数据。

RLS 采用了分布式注册，这意味着它在不同的地点可能包括多个服务器。通过分发这些 **RLS** 注册，可以增加系统的整体规模并且储存比单一系统更多的映射。如果需要，**RLS** 也可以被部署成为一个单一的、集中模式的服务器。实例如下：

使用 **RLS** 作为部分数据管理框架的系统有 **Laser Interferometer Gravitational Wave Observatory (LIGWO)** 项目。**LIGWO** 的科学家在两个地点放置设备来探测重力波的存在。在一次科学实验过程中，每台 **LIGWO** 设备会产生数百万个数据文件。8 个其他地点的科学家需要拷贝这些数据集到他们本地的存储系统中，以便对这些数据进行科学分析。因此，每个 **LIGWO** 数据文件都可能在网格中不同的地点被复制了 10 次。**LIGWO** 在每个地点都部署了 **RLS** 服务器来注册本地映射并且收集有关其他 **LIGWO** 地点映射的信息。为了寻找一个数据文件的复制品，科学家向 **LIGWO** 的数据管理系统发出请求，之后调用 **Lightweight Data Replicator (LDR)**。**LDR** 查询 **RLS** 发现本地文件拷贝，如果没有，**RLS** 就告诉数据管理系统这个文件在网格上的位置。之后 **LDR** 系统产生一个将该文件复制到本地的存储系统并且将其注册到本地 **RLS** 服务器中的请求。

LIGO 目前使用 **RLS** 来进行数据管理，这个系统在超过 300 万的逻辑文件和 3 千万个物理文件之间建立了映射。

（3）运行管理

① 社区调度框架（Community Scheduler Framework, CSF）

社区调度框架（**CSF**）4.0 是一个基于 **WSRF** 的网格级元时序安排框架，建立在 **Globus Toolkit** 基础上。**CSF** 为网格用户提供了接口和工具来提交工作，创建高级预约并且在网格级定义不同的时序安排策略。通过使用 **CSF**，网格用户可以通过单一接口访问不同的资源管理器，比如 **LSF**, **PBS**, **Condor** 和 **SGE**。目前已经测试过的运行平台为 **Linux (RedHat 9.0)**。

CSF支持的其他功能特性：在网格级提交、控制和监视作业情况，在网格级创建

和管理高级预约，发送作业与高级预约操作给本地资源经理，提供一个时间调度接口的插件，通过这个插件实现网络和用于具体的时间调度策略，创建工作队列，每个都有分别定义好的时间调度策略。CSF主要依赖于以下的GT组件或第三方软件组件：Pre-WS Authentication and Authorization, OpenSSL, Java WS Core, Message-Level Security, Delegation Service, GRAM, MDS - internal libraries, LSF, 一些其他被GRAM接口支持的批量时间调度器。

② 网络资源分配管理（Grid Resource Allocation Manager, GRAM）

Globus Toolkit 的 GRAM 组件提供了一组 Web 服务，它们采用了 WS-Resource Framework，其设计目标是帮助用户在网格环境中提交作业和管理作业。GRAM 可以用于那些需要信任证书、可靠执行以及协调文件步骤的作业。GRAM 可以帮助实现信任证书的管理，从而提交作业，监视作业的进展状况，控制作业的执行情况，并分阶段地处理相关的数据。不同的信任证书可以用来代表不同的功能。GRAM 提供了一个可靠的执行环境，并集成了多个专门用来在复杂环境中对作业执行管理进行优化的调度器。GRAM 和支持服务提供了文件分段传输的功能，甚至还提供了在作业运行时访问输出数据文件的功能，从而有效地实现对输出数据的利用。

③ GRAM 调度器

当用户提交一个作业时，GRAM 创建一个 ManagedJob 对象，并返回其句柄。然后用户就可以使用这个句柄与这个作业进行交互，这实际上是一个 网络服务寻址终端参考（WS-Addressing Endpoint Reference, EPR）。

网络作业可能会消耗大量的计算周期，因此一个网格环境通常会包含很多资源集群。这些集群通常由一个调度器进行控制。调度器负责对资源的使用进行优化，按照定义好的策略有效地执行作业。有很多调度器可以使用，其中包括 Load Sharing Facility、Condor 和 Portable Batch System。

GRAM 使用调度适配器将作业描述语法转换成调度器可以执行的命令。除了为调度器提供一个接口之外，GRAM 还包含了一种新的 UNIX® 进程来启动作业的机制。这适用于小型的串行作业，它们在运行时不需要集群资源，也不需要共享内存或消息传递。

派生进程使用一个 Fork Starter 程序来启动并监视针对 GRAM 的作业进程。它还记录了一些数据，例如启动时间、结束时间以及该作业的退出代码。当提交作业时，事件会被写入日志文件中。当有重要的事件发生时，会继续在后面添加日志。这些调度器日志文件由称为 Scheduler Event Generator（SEG）的进程进行监视，它会让 GRAM 通过监听相关的事件并与其通信，来保持与作业状态的同步。

GRAM 提供了作业和执行管理服务来提交、监视和控制作业，但是这都依赖于用来传输文件和管理证书所使用的支持服务。GridFTP 所提供的文件服务用来帮助 GRAM 实现对输入和输出文件的分段传输。证书管理服务处理证书对其他服务和所需要的分布式网格资源的委托。

④ Pre-WS GRAM

GRAM 服务提供了一个单一的接口来获取或者使用远程系统资源类执行作业。GRAM 的最常见用途是远程应用执行和控制。它被设计用来为时间调度系统提供一个统一的、可扩展的接口。已经测试过的系统包括 Linux。

Pre-WS GRAM 提供了其他特性：远程作业执行与管理，对批处理时间调度系统的统一的可扩展的接口，作业执行之前和之后的文件存储，作业执行过程中的数据流标准输出/错误比值。Pre-WS GRAM 主要依赖于以下的 GT 组件或第三方软件组件：C 通用库（Common Libraries），（Pre-WS Authentication and Authorization 认证与权限），XIO（extensible input/output library，可扩展输入输出库），GridFTP，PBS（工作负载管理器和调度器），Condor，LSF（Load Sharing Facility，负载共享设备），其他的批处理时间调度器。

（4）信息服务

① 监测和发现系统（Monitoring & Discovery System，MDS4）

监测和发现系统共包含了以下 4 个模块：索引服务（Index Service）、触发器服务（Trigger Service）、聚合器框架（Aggregator Framework）、Web 监测和发现系统（WebMDS）。

② 索引服务（Index Service）

索引服务从网格资源中收集监测和发现信息，并且将其发布到一个单一地点。总的来说，最好的情况是一个虚拟组织在虚拟组织中部署一个或者多个索引服务，这些服务将会从所有的有效网格资源中收集数据。目前已经测试过的运行平台有 Linux on i386、Windows XP，已经测试过的运行容器有 Java WS Core container、Tomcat 5.0.28。

在 GT4 中 Index Service 的新特性：基于 WSRF 而不是 OGSi，聚合数据通过 WS-Resource 属性机制发布，聚合器框架模块使用 WS-Resource 属性收集机制从被监控资源收集数据，包括基础的 WS-Resource Properties poll 操作和 WS-Notification，聚合的持久配置被重构。索引服务需要的技术支持：Java WS Core，WEB 监测和发现系统，聚合器框架。

③ 触发器服务（Trigger Service）

触发器服务从网格资源中收集数据。如果管理员定义了匹配规则，那么服务可以执行多种行为。一个典型的应用是，当计算机资源中的队列长度超过一个临界值时，就会发送 email 提醒。目前已经测试过的运行平台有 Linux on i386、Windows XP，已经测试过的运行容器有 Java WS Core container、Tomcat 5.0.28。

触发器服务支持的其他特性：监控资源属性来匹配触发条件，当一个触发条件被匹配的时候，就执行一个定制行为，比如，给管理员发送 E-mail。被监控的服务通过基于分组注册 API 的服务来管理，允许相同的许多用户使用一个索引服务。触发器服务需要的技术支持：Java WS Core，Web 监测和发现系统，聚合器框架。

④ 聚合器框架 (Aggregator Framework)

聚合器框架是一个软件框架，Web 监测和发现系统服务基于它而构建。聚合器框架从一个聚合器源收集数据并且把这些数据发送给一个用于处理的聚合器槽。通过 Globus Toolkit 分发的聚合器资源包括查询服务模块，收集数据模块以及执行程序产生数据模块。聚合器槽包含实现 Web 监测和发现系统索引服务接口和 WEB 监测和发现系统触发器服务接口的模块。目前已经测试过的运行平台有 Linux on i386、Windows XP。

GT4 中聚合器框架支持的新特性：开始使用 WSRF 机制（之前使用的是 OGSi），通过执行本地脚本来收集信息的额外资源，APIs 聚合管理通过 WS ServiceGroup APIs 来执行。聚合器框架支持的其他特性：通过使用可插入的聚合资源从网格资源中收集信息，把收集到的信息传输给可插入的信息槽，管理个人聚合注册的创建和销毁。Aggregator Framework 需要的技术支持：Java WS Core。

⑤ Web 监测和发现系统 (WebMDS)

Web 监测和发现系统是 GT4 中的一个新组件，它使终端用户无须安装任何其他软件，就可通过一个标准的 web 浏览器接口来浏览监控信息。Web 监测和发现系统通过 servlet 来实现，使用了插入接口来收集监控信息和 XSLT 格式转换，并且将数据以一种可读格式呈现给用户。Web 站点的管理员可以通过使用 HTML 表格选项来定制他们的 WebMDS 部署，并且配置不同的插件来收集数据或者实现 XSLT 格式转换，并且创建他们自己的插件或者 XSLT 格式转换。目前已经测试过的运行平台有 RedHat Linux (i386)、Windows XP。测试过的容器有 Tomcat version 5.0.28。

GT4 中 WebMDS 支持的新特性：支持多种机制来收集监控信息和 XSLT 格式转换的可扩展的插件接口，通过资源属性机制获取监控信息的插件，通过读取本地文件来实现 XSLT 格式转换的插件。WebMDS 需要的技术支持：Java WS Core, Tomcat。

⑥ Pre-WS MDS (Monitoring & Discovery System2, MDS2)

Pre-WS MDS 实质上还是和 Globus Toolkit 2.4 中的基于 LDAP 和 GRIS/GIIS 协议的 MDS 2.4 一样，Pre-WS MDS 提供了一种标准机制来发布和查找资源状态和配置信息，它为低层次数据提供者收集的数据提供了一种统一的、灵活的接口。它有一个可扩展的分散结构，并且可以处理静态或者动态数据。

检测与发现系统需要的技术支持：OpenLDAP 2.0.22、OpenSSL 0.9.6b 和 SASL cyrus-sasl 1.5.27。

(5) 常用运行组件

① Python WS Core

作为一种 Python WS Core, pyGridWare 提供了一个用于创建基于 WSRF 的 Web 服务的基础 Python 工具集，已经证明可以和 Java WSRF 互操作。性能是目前主要的考虑和动机。这个 WSRF 项目通过对性能要求十分严格的安全和 DOM 层使用 C 的实现，大大增加了系统效率。

WSRF support 包括 WS-Resource Lifetime , WS-Resource Properties 和 WS-Notification。同时还包括升级后的 WS-Secure Conversation 和 Secure Message 的支持。

Python WS Core需要的技术支持有: Python 2.3, pyXML-0.8.4, 4Suite-1.0a4, Twisted-1.3.0, WS-Security: XML 数字签名, Secure Conversation等, pyGlobus Python Wrapper for GT2, Globus 2 Security and utility packages。

② C WS Core

C WS Core 提供了一个 C 的基础工具集来创建基于 WSRF 的 Web 服务和用户, 它们都遵循 WS-Resource 和 WS-Notification 详细规范。测试过的系统平台有: IA32/Linux/gcc32 、 IA64/Linux/gcc64 、 x86_64/Linux/gcc64 、 SPARC/Solaris 9/vendorcc32、PowerPC/AIX 5.2/vendorcc32、Mac/OS X/gcc32。

C WS Core支持的特性: 绑定产生, 直接来自WSDL模式的绑定产生, ANSI-C 句柄和纲要, 用于撰写事件驱动代码的非阻塞的用户句柄, EPR (EndpointReference) 封装, WSRF支持的客户句柄和服务, HTTP/1.1 的支持, 可嵌入的服务API, 独立的服务容器, 基于WSRF的服务。C WS Core需要的技术支持: C Common Libraries, Pre-WS Authentication and Authorization (GSI) , Globus XIO, Libxml2 , OpenSSL, JavaScript。

③ Java WS Core

Java WS Core是一个 WSRF和WSN (WebService Notification, Web服务通知) 家族标准的实现。它提供了用于创建有状态Web服务的API和工具。Java WS Core 需要 J2SE 1.3.1 以上版本支持, 已经测试过的运行平台包括Linux (Red Hat 7.3)、Windows 2000 、 Windows XP、Solaris 9。测试过的JVMs 包括: Sun JVM 1.3.1, 1.4.2 和 1.5.0、IBM JVM 1.3.1, 1.4.1, 和 1.4.2、BEA JRockit JVM 1.5.0, 这里需要注意的是: 不支持 GCJ (GNU Compiler for Java)。测试过的容器包括: Tomcat 4.1.31、Tomcat 5.0.30。

GT4 中, Java WS Core支持的新特性: 2004/06 OASIS WSRF和 WSN 工作草案详细规范的实现, 基本 HTTP/1.1 客户端&服务器支持, 基于Apache Tomcat的JNDI 服务注册, Work Manager and Timer 详细规范的实现。Java WS Core支持的其他特性: 一个独立的可嵌入式容器, Tomcat 4.1 和 5.0 的支持, 资源保持和修复的基本API, 长久签署的支持, 自动服务和 ResourceHome的开机激活, 操作执行。Java WS Core需要的技术支持: Java CoG Kit, Apache Xerces, Apache XML Security, Apache Axis, Apache Xalan, Apache Addressing, Apache Commons BeanUtils, Apache Commons CLI, Apache Commons Collections, Apache Commons Digester, Concurrent Library, Apache Tomcat JNDI。

④ C Common Libraries

C Common Libraries 提供了一个数据类型的抽象层, 包括 libc 系统调用以及在 GlobusToolkit 中会用到的数据结构或应用。

C Common Libraries 在 GT4 中包含的新特性：增加了 globus_range_list abstraction，增加了 globus_logging abstraction。C Common Libraries 需要的技术支持：globus_core module。

⑤ XIO

Globus XIO 是一个为 Globus Toolkit 用 C 语言编写的可扩展的输入/输出类库。它提供了一个支持多线程协议的 API（打开、关闭、读写功能），同时封装了实现协议作为驱动。GT4 中包括的 XIO 驱动包括 TCP, UDP, file, HTTP, GSI, GSSAPI_FTP, TELNET 和 queuing。

除此之外，Globus XIO 为协议开发人员提供了一个驱动开发接口。这个接口允许开发者集中精力于协议代码的开发而不是构架，比如 XIO 提供了一个错误处理框架，异步消息传输也超时处理等。

XIO的基于驱动的方法，通过支持驱动堆栈使代码重用程度最大化。XIO驱动可以被编写成为原子单位并且压入堆栈中。这个模块化设计提供了最大的灵活性并且简化了设计和每个协议的评估过程。测试过的运行平台包括：Linux、Mandrakelinux release 10.1、SuSE Linux 9.1 (i586)、Debian GNU/Linux 3.1、Red Hat Linux release 9、SunOS、SunOS 5.9 sun4u sparc SUNW,Sun-Fire-280R、MacOS、Darwin Kernel Version 7.9.0、Additionally all platforms supported by GT4.0 Platforms。

GT4 中，XIO 支持的新特性：UDT 驱动，Mode E 驱动，Telnet 驱动，Queuing 驱动，Ordering 驱动，Dynamically loadable 驱动。XIO 支持的其他特性：单一 API 的可交换 IO 实现，异步 IO 支持，本地超时支持，提供驱动详细描述的数据描述符，用于使代码重用最大化的模块驱动堆栈，TCP, UDP, file, HTTP, telnet, mode E, GSI 驱动。XIO 需要的技术支持：Globus Core, Globus Common, Globus GSSAPI。

6.5 发展 趋 势

网络技术发展的主要趋势有如下几方面。

(1) 标准化趋势

标准化的网络协议有助于网络实现共享和互通。迄今为止，网络计算还没有正式的标准，但在核心技术上，相关机构与企业已达成共识：Globus Toolkit 已成为网络计算事实上的标准。

(2) 技术融合趋势

在 OGSA 出现之前，已经出现很多种用于分布式计算的技术和产品，各种以填补异构平台之间的差异为己任的网格平台，如 Condor、Legion、Globus 等，但这些平台也是互不兼容的。Web Service 技术的出现有望解决这一局面，它在各种异构平台之上构筑了一层通用的、与平台无关的信息和服务交换设施，从而屏蔽了互联网

中千差万别的差异，使信息和服务畅通无阻地在计算机之间流动。它为工业应用提供一种互联互通的标准。只有充分利用 Web 服务技术，才能真正使得网格成为未来的网络基础设施。这在 OGSA、WSRF 中得到体现。

（3）网格和 Web 服务相融合

作为 OGSA 最新核心规范的 Web 服务资源框架 WSRF，它的提出加速了网格和 Web 服务的融合，以及科研界和工业界的接轨。对 WSRF 本身而言，由于其提出不久，其规范还有待在实践中进一步应用证明，并逐步完善。

第 7 章 科研数据的采集、管理、保存与分析技术

信息技术的加速发展正推动着科学研究的高速发展，伴随着数据与信息爆炸，计算能力和网络带宽的需求也随之而巨大增长。

数据被认为是 e-Science 的灵魂。如何对科学数据进行采集、管理、保存和分析是一个非常重要的课题。传统的科学研究只能通过人将实验、数据、计算机和分析设备连接在一起，无法得到有效的技术支持与保证。随着网络功能的日臻完善，不同的服务得以便利地连接在一起，这使得现代科学研究得以摆脱传统方式的束缚，借助网络来实现数据和信息的高速传输与计算。于是，如何从试验室、观测中心和感应器中接收研究数据，并对这些研究数据进行高质量的管理、保存、分析，使其长期可靠使用逐渐成为摆在科学家们面前的一个重要的问题。

本章将对在 e-Science 环境中对科学数据进行采集、存储、集成、管理、保存和分析的关键技术进行分析。

7.1 数据采集技术

科学研究有各种专用的设备，例如粒子加速器、天文望远镜、同步辐射装置、各种传感器等等，这些设备可以帮助科学家实现对科学数据的采集。这些科研设备采集到的科学数据是 e-Science 研究工作的基础。

本节主要分析 3 种在 e-Science 环境中进行数据采集的重要技术，并通过实例分析在 e-Science 环境中，数据的采集如何实现。

1. 传感技术

传感技术同计算机技术、通信技术一起被称为信息技术的三大支柱。从仿生学观点，如果把计算机看成处理和识别信息的“大脑”，把通信系统看成传递信息的“神经系统”的话，那么传感器就是“感觉器官”。

传感技术是从自然信源获取信息，并对之进行处理（变换）和识别的一门多学科交叉的现代科学与工程技术，它涉及传感器（又称换能器）、信息处理和识别的规划设计、开发、制/建造、测试、应用及评价改进等活动。获取信息靠各类传感器，它们有各种物理量、化学量或生物量的传感器。按照信息论的凸性定理，传感器的功能与品质决定了传感系统获取自然信息的信息量和信息质量，是高品质传感技术

系统的构造第一个关键。信息处理包括信号的预处理、后置处理、特征提取与选择等。识别的主要任务是对经过处理信息进行辨识与分类。它利用被识别（或诊断）对象与特征信息间的关联关系模型对输入的特征信息集进行辨识、比较、分类和判断。因此，传感技术是遵循信息论和系统论的。它包含了众多的高新技术、被众多的产业广泛采用。它也是现代科学技术发展的基础条件，应该受到充分地重视。

在 e-Science 研究中，SensorWeb（传感器 Web）的概念得到了广泛的关注。SensorWeb 是 NASA 科学家提出一个概念。一个 SensorWeb 的框架本质上是一个具有一定自适应能力的框架，在 SensorWeb 中，它的传感器，计算资源和存储资源是根据具体情况能够进行相互调整的，它可以动态地修改、适应它们的测量模式、观测策略、以及处理状态，能够智能地收集、交换和综合传感器数据和其他信息，获得最大的科学回报。SensorWeb 的框架如图 7-1 所示。

SensorWeb 的概念在地球科学、空间物理学等等方面都得到了充分的利用。NASA 认为 SensorWeb 是一种新型的环境监测和探索设施。是一种新型的地理信息系统（GIS），它可以嵌入到环境之中，进行监测和控制，作为空间分布的同步设备，能够适应变化的环境。

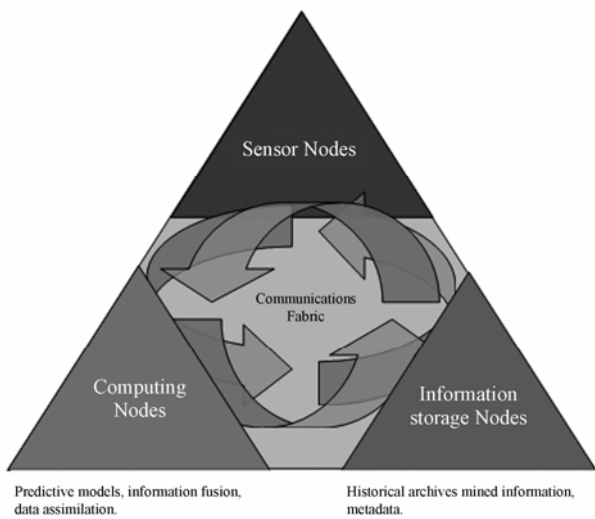


图 7-1 SensorWeb¹

2. 遥感技术

遥感技术包括传感器技术，信息传输技术，信息处理、提取和应用技术，目标

¹ Stephen J. Talabac, Sensor Webs:An Emerging Concept for Future Earth Observing Systems,[EB/OL]. (2008-8-20), http://eos.gsfc.nasa.gov/eos-ll/docs/sensor-webs_Talabac.ppt

信息特征的分析与测量技术等。遥感技术依其遥感仪器所选用的波谱性质可分为：电磁波遥感技术，声纳遥感技术，物理场（如重力和磁力场）遥感技术。电磁波遥感技术是利用各种物体/物质反射或发射出不同特性的电磁波进行遥感的，可分为可见光、红外、微波等遥感技术。按照感测目标的能源作用可分为：主动式遥感技术和被动式遥感技术。按照记录信息的表现形式可分为：图像方式和非图像方式。按照遥感器使用的平台可分为：航天遥感技术，航空遥感技术、地面遥感技术。按照遥感的应用领域可分为：地球资源遥感技术，环境遥感技术，气象遥感技术，海洋遥感技术等。遥感技术在区域地质调查、生态环境调查、矿产资源调查等等领域都有着重要作用。

3. 宇宙射线探测技术

宇宙射线观测属于基础性研究，对探索大气物理、核物理、太阳活动、无线通信、河外星系伽马源、高能粒子特性、天气变化等有着重要意义，对于揭示物质和能源的奥秘起到重要作用。

4. DiscoveryNet的传感器网络

DiscoveryNet 是一个基于 Grid 的项目，对来源于生命科学、环境科学和遥感等领域的，各种分布式高吞吐量设备所产生的数据进行集成和分析的项目。它的目标是开发一个先进的计算基础设施，支持实时地对一些时效性很强的大容量数据的处理、解析、集成、可视化和精简。DiscoveryNet 项目中的一个应用领域是对 GUSTO（Generic Ultraviolet Sensors Technologies and Observations，高吞吐量污染监测传感器）所产生的数据进行分析。

在目标区域（如工业化程度较高的地区和人员密度较大的地区）构建一个传感器网络，并收集大量的数据有助于通过分析发现新的规律。这些分析包括对各种污染物在不同时空地区出现情况的分析和预测，以及这些污染物与第三方数据如天气、健康和交通的数据的相互关系。这些分析结果可以帮助研究人员发现什么对当地的人员健康构成威胁。

然而，每个 GUSTO 高吞吐量污染监测传感器每天能够产生 8GB 的数据，这对于如何管理和分析所收集的数据提出了一个严重的问题。针对这一问题，DiscoveryNet 提出了 Sensor Grid（传感器网络）的解决方案，它需要解决的问题包括：分布式的传感器数据存取和集成；大型的数据集存储和管理；分布式的引用数据存取和集成；高强度的开放的分析计算。

DiscoveryNet 的 Sensor Grid 框架如图 7-2 所示。

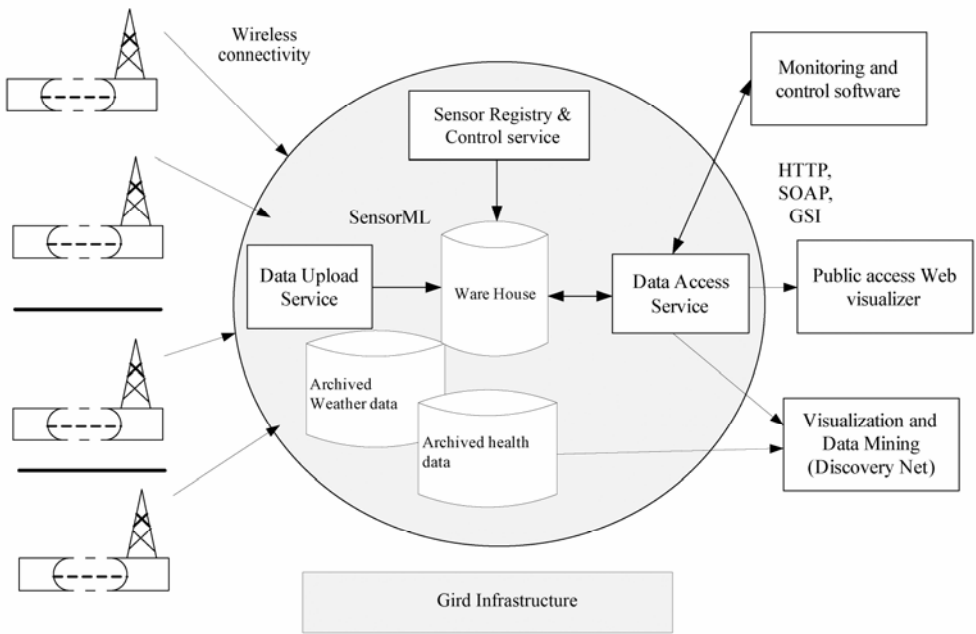


图 7-2 Sensor Grid框架¹

在每个 GUSTO 高吞吐量污染监测传感器中，安装了计算机，它能够分析传感器读取到的数据，并且每两秒生成一个对污染物的测量数据，并将其提交到远程服务之上，进行数据仓储。

7.2 数据集成技术

由于现实情况下，科学数据基本上是各个研究机构分散地采集和收藏的。e-Science 中需要解决的一个问题就是如何通过网络基础设施将分散的科学数据集成到一起，形成可共享的科学数据。在 e-Science 研究领域，数据集成的相关技术研究主要包括：英国国家 e-Science 中心对 OGSA-DAI 新型框架的探索、BRIDGES 对 Information Integrater 的应用等。

7.2.1 新型OGSA-DAI框架

开放网格服务框架——数据访问和整合 (the Open Grid Services Architecture—

¹ Moustafa Ghanem, Distributed Data Mining in Discovery Net,[EB/OL].(2008-8-20), [http:// www.nesc.ac.uk / action/esi/download.cfm?index=2041](http://www.nesc.ac.uk/action/esi/download.cfm?index=2041)

Data Access and Integration, OGSA-DAI) 项目就是为提供一种统一的数据资源访问模式而建立的。它的目的是让客户端能够向远端服务器提交自己的需求文档，并以此来完成自己的一系列任务。到现在为止，OGSA-DAI 项目主要致力于开发一个可扩展的基于文档框架，以提供一个统一的访问大量不同数据资源的接口。虽然这种思路与当前的 DAIS (Data Access and Integration Services, 数据访问与集成服务) 发展特点不尽相同，但是它们两者之间并没有很实质的区别，其主要区别仅仅在于提交服务的请求的尺度而已。在某种情况下，未来的 OGSA-DAI 会采用现在所定义的 DAIS 接口，尽管这种基于文档的访问方法在将来还是会收到很大的支持。

英国国家 e-Science 中心认为基础的数据集成能力在当前的框架中并没有得到很好的支持，不过利用现有的核心 OGSA-DAI 分布还是可以勉强做到这一点的。然而，这就依赖于一个 OGSA-DAI 客户端是否可以明确地将服务联合在一起。很明显，这种方法很难控制，同时不容易实施。英国高性能计算研究机构 EPCC 已经开始进行简单的数据整合，希望以此能够理解现有的 OGSA-DAI 分布状况对数据整合的支持程度，并且确定扩展数据整合能力所需的附加功能。OGSA-DAI 并不需要提供整个的数据整合层，可以对数据资源自己提供的整合能力进行一次开发，或者通过第三方的方法，将这些已经被 OGSA-DAI 同其他类型的数据资源联合在一块的数据资源进行包装（见图 7-3）。



Key to Symbols			
Data Flow			
→	Call	--→	Response
Actors			
	Non-OGSI process		OGSI process
A	Analyst	G	GDS
C	Consumer	P	Producer
Data			
Q	Query definition	D	Delivery definition
S	Status	R	Result
U	Update data	I	Data and (URL)

图 7-3 数据集成¹

¹ Mario Antonioletti etc. OGSA-DAI Usage Scenarios and Behaviour: Determining good practice [EB/OL].(2008-8-20), <http://www.allhands.org.uk/2004/proceedings/papers/266.pdf>

更加完善的数据整合能力是由分布式查询进程（Distributed Query Processing, DQP）实现的，目前作为一个单独的包放在 OGSA-DAI 网站上。DQP 是作为一个抽象层位于 GSA-DAI 之上，其功能是将用户提交的查询应用到不同的数据资源上去，就像是本地资源一样。实现这种功能的方法是增加了一套网格服务，扩展了 OGSA-DAI 的范围。这些服务其中的一个作为一个客户端的联系点，并且将幕后其他的协调服务协调在一块，其中也包括每个数据资源上评估查询内容的服务。数据整合过程在客户端或者服务端都可以进行，DQP 作为一个 OGSA-SAI 的一个扩展在服务端运行，允许数据整合的运行。项目未来的工作就是研究是否可以将这种功能直接整合到 OGSA-SAI 的核心分布中去（图 7-4）。

图 7-4 数据访问整合过程¹

早期 OGSA-DAI 的设计是基于这样一个假设：web 服务是轻量级的。因此，OGSA-DAI 就需要依靠 factory 模式来建立一个能够与正在处理的数据有一对一关系的独立的网格服务集合。这意味着网格服务句柄必须提供一个数据命名系统，从而使得错误对于一个特殊的对话来说只是局部的，而且只有访问和处理过的数据才有持续提供的需求。

最初的 OGSA-DAI 是建立在 GT3 提供的 OGSF 架构上的, 此后, OGSA-DAI 已经发布了基于 Axis、WS-I、OMII 1、WS-I 和 GT4 上的版本了。它提供了对相关的 XML

¹ Mario Antonioletti etc. OGSA-DAI Usage Scenarios and Behaviour: Determining good practice [EB/OL].(2008-8-20), <http://www.allhands.org.uk/2004/proceedings/papers/266.pdf>

数据库和索引文件的直接访问，还有数据翻译和第三方传递服务。这种行为框架为应用开发者们增加自己的行为提供了可能。OGSA 就为 WS-I 的发展提供了一个参考框架。从 OGSIdao WS-I+和 WSRF 的转换，需要在架构上进行一次改变，这是因为 Web 服务现在必须能够适应长期运行。这样，就必须对 OGSA-DAI 的方方面面重新考虑，而且还要考虑到扩展运用而可能带来的性能和操作问题。

于是，在新的 OGSA-DAI 架构中必须考虑以下的一些问题：以前所有的行为和工具都必须继续支持；每个数据服务必须支持零个或者更多的数据源，必须可以对数据资源进行动态的设置；对复杂命名系统有了更多的需求；每个数据资源都是可以设置的，在操作时也可以重新设置；每个数据服务都应该可以管理，而且是持续存在的；同样的架构应该对小型任务和有大量数据的任务有相同的请求；数据服务应该支持并发对话和事务；请求文档的作用应该更大，一概可以支持更多的数据源和操作；数据服务应该能够容纳标准接口；数据资源应该是一种 OGSA-DAI 的扩展，以便可以动态的展示虚拟数据资源；基于行为的扩展框架应该能够保护应用开发者的投资。

当前的 OGSA-DAI 架构如图 7-5 所示。

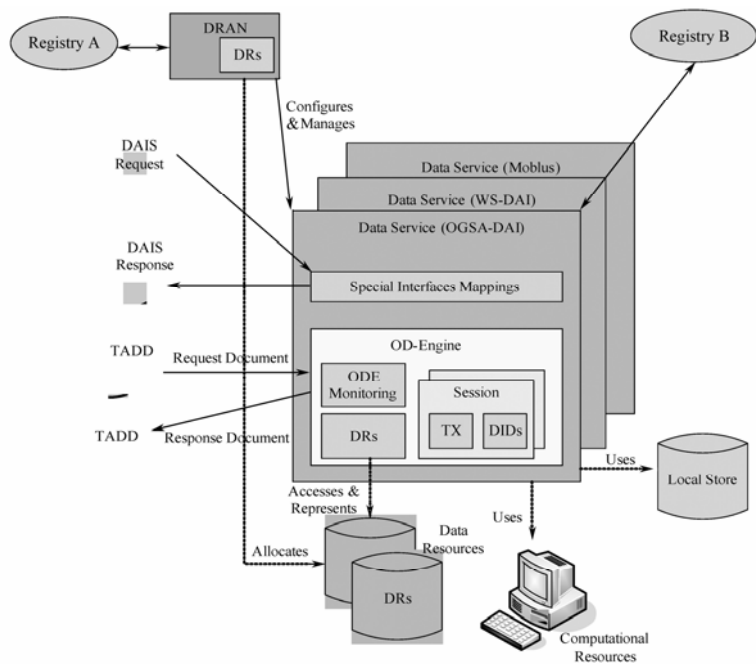


图 7-5 OGSA-DAI架构¹

¹ Atkinson, M. A new Architecture for OGSA-DAI,[EB/OL].(2008-8-20), <http://www.ogsadai.org.uk/documentation/publications/NewArchitecturePaperAHMSep2005.pdf>

(1) 数据源 (Data Source)

这里提到的数据资源指的是任何能够作为一种数据源或者数据容器的实体，并且拥有自己的管理框架。OGSA-DAI 提供了一种可扩展的框架，可以用一种灵活的方式来表现一个数据源。数据源并不一定要是一个物理资源，但是必须能够提供复杂的虚拟数据资源。

(2) 数据服务 (Data Services, DS)

数据服务为客户端提供了接入点。它可以在注册时或者利用 DRAM 被发现。数据服务可以呈现零个或者多个数据资源。每个数据资源可以被动态的设置。OGSA-DAI 的设计目的就是为了支持多种不同的平台。对于每个资源，客户端都需要提交一个对于 DS 来说正确的请求。OGSA-DAI 提供了一个客户端的类库 CTK (Client Type Knowledge)，它提供了一个可编程的抽象和一个所有平台的通用接口。

(3) 行为与任务 (Activity and Jobs)

由于数据资源行列的不断扩大，有必要有一个开放的方案。通过这个方案，能够添加数据资源，描述这些资源所支持的语言和操作。在 OGSA-DAI 中，定义了“行为”来实现这个目的。

在 OGSA-DAI 中，一个行为代表一个基本的工作元素。一个任务就是在某个特定数据上，按照某个特定的参数来执行这个行为的过程。行为可以达到各种各样的功能，大致分为 3 个主要的功能组：① 声明，提供了对数据资源的访问；② 翻译，将数据进行转型；③ 传递，从第三方获得数据或者向第三方提供数据。

OGSA-DAI 提供了一种可扩展的行为框架。开发者可以根据自己的需要轻松地添加行为。这种扩展很灵活，并且很有价值。

(4) 任务和需求文档 (TADD)

TADD (Task and Data Document，任务和数据文档) 支持对多任务的定制，包括这些任务的参数、输入、输出、任务之间的数据流和控制流。这个复合的格式可以用于提交简单的请求，避免了不必要的延迟，可以报告结果状态，可以代表或者级联工作。其容纳数据的能力进一步减少了以往一些必须产生的交流所带来的成本。

控制流允许客户端来定制他所要执行的任务的顺序，哪些任务可以并行处理。数据流支持行为之间的通信，并以行为之间的通信流来建立有效的数据处理管线。

(5) 会话 (Session)

对话支持几个 TADD 之间的状态 (上下文) 的共享和管理。对话是 OGSA-DAI 内部的实体。对于一个加入对话的请求，TADD 会保存正确的信息。

(6) 事务 (Transaction)

对于很多技术来讲，为了确保资源的完整性，事务都是很重要的。两个最主要的应用就是处理潜在的由并发访问和执行事务时系统崩溃或者出错后进行错误恢复时所带来的不一致性。一个事务会将很多行为组合到一个执行体当中，而且一个单独的行为还有可能包含多个底层操作。这些都可能产生影响，事物状态会被存储在

一个对话当中。一个事务的处理过程会被保存在一个事务记录中，以便在出错后恢复组件在恢复数据资源，将潜在的资源恢复到一个一致状态时使用。

当一个 DS 代表了多个资源，并且 TADD 包含一个对多个资源的事务请求时，数据资源中的事务协调组件支持 DS 内部的分布式事务。而且，可以对多 DS 进行分布式事务处理。

(7) 数据标识符 (Data ID)

数据标识符用来描述内部或者外部的结果处理。用标识符标识过的数据包含了这个数据的初始信息、数据格式，当然还有数据本身。数据标识符保存在一个对话当中，这样一个请求中的其他任务或者这个对话中的子任务都可以访问这个 ID。

(8) OGSA-DAI 引擎 (OGSA-DAI engine)

DS 的核心是 OGSA-DAI 引擎 (ODE)，它提供了一个行为和任务管理框架。ODE 用来支持动态的设置、对话、事务、恢复和并发。

(9) 特殊接口

在新的架构中，一个重要的需求就是 DS 要包含标准的 Web 服务接口，尤其是 WS-DAI，或者那些需要特殊信息协议的接口，比如分布式事务协调协议。将这些接口的端口绑定到建立相等的执行图的代码上，就可以调用这些特殊接口。传输组器会将所得到的结果以 TADD 的格式输出，接着转变成所需要的结果格式。

(10) 数据资源访问管理器 (Data Resource Access Manager, DRAM)

DRAM 是一个持久的服务，提供对一套数据服务的管理和操作接口。它可以初始化和设置一个 DS，同时也可以监视和控制它。

(11) 命名 (Naming)

命名的作用是标识一个实体。

另外还有很多其他的基于 OGSA-DAI 的数据访问和整合解决方案。

OGSA-WebDB，它提供了一个基于 OGSA-DAI R5 接口的网格数据服务接口，已集成现有的网络数据库到网格环境下。这个系统由一个代理数据库，一个调解器和一些数据库包装器组成。该系统允许用户可以通过一个两步的处理计划对多个网络数据库进行分布式 SQL 查询。在这个两步的处理计划中，一个 SQL 语句会被处理两次：一次由调解器，一次由 OGSA-DAI。

调解器首先针对每个网络数据库将 SQL 查询转换成面向网址的布尔查询条件，然后调用面向网址的包装器来执行这个查询。一旦所有的结果被装入代表每个网络数据库的个人代理表单后，OGSA-DAI 就会第二次对这个个人代理表单运行该 SQL 查询以便将结果整合在一起。这个内部数据和外部数据之间的映射是在代理设置过程中定义的。

网格数据调解服务 (GRID Data Mediation Service, GDMS)。它提供了一个更通用的网格数据集成方法。与 OGSA-DQP 不同，它的主要目的是调整数据资源的异构性。该服务支持内连接和联合，可以和关系、XML 和 CSV 等文件资源同时使用，

但是不能和网络数据库同时使用。DGMS 有一个调解—包装架构，这个架构可以将多个数据源按照一个全局的框架联合起来，然后用一个单独的虚拟 OGSA-DAI R3 网格数据资源来标识他们。这个全局框架和所关联的数据源之间的关系是按照一种静态设置的映射框架来定义的。客户端可以用以全局框架表达的 SQL 语言以一种标准方式对网格数据资源进行查询。GDMS 利用这个映射框架来重新表示这个查询，并且将这个查询分成块，接着通过定制查询包装器一块一块地执行这个查询。

7.2.2 BRIDGES ——基于OGSA-DAI进行信息集成的实例

BRIDGES (Biomedical Research Informatics Delivered by Grid Enabled Services, 基于网格服务的生物化学信息学)¹由DTI (Department of Technology and Industry, 英国技术工业部) 所资助的, 目的是辅助CFG (Cardiovascular Functional Genomics, 心血管功能新基因组学) 项目区建立一个网格架构, 提供对数据的相关的、快速的、可靠的和用户友好的访问。

当前在世界各地还分布着很多的包含生物信息的数据库, 大多数的主要染色体组数据库都有浏览器接口, 通过这个接口用户可以输入一系列的参数, 从而得到很多更详细的信息, 或者说一个特定的基因序列。这经常会让科学家一步一步地按照这些网络资源给出的超链接选项走下去, 直到科学家最终找到所需要的信息, 或者让科学家参考另外一个网页以获得详细的信息。很明显, 这个过程耗时很多, 而且把所有的相关数据整合到一个文档中也是比较困难。结果科学家们不得不在这种即耗时又不智能化的环境中进行检索研究, 代表现象就是在不同的生命科学数据源中“跳跃”。

网络技术为克服这种困难提供了可行性。通过一个查询就可以对多个集成在一块的数据库同时进行检索, 而且返回的结果也可都集成在一块。BRIDGES 就为此使用了两个主要技术: OGSA-DAI 和 IBM 的数据集成器。CFG 项目的目的是研究造成高血压的可能的遗传因素。这个项目得到了五个英国部门和一个荷兰部门的共同参与。BRIDGES 就是要为 CFG 提供一个基于网格的计算和数据架构来满足 CFG 的需要。

BRIDGES 中一个关键组件就是数据集线器 (Data Hub) (图 7-6)。这个不但代表了一个本地的数据仓库, 同时还代表了通过外部网格数据集所共享的数据。这些数据集存在于很不同的远端地点, 并且也有着很不同的安全需求。一些数据是公共的, 而其他只能由特定的 CFG 项目合作伙伴使用。

信息整合器本质上就是一套对那些相关或者不相关的, 能够扩展 DB2 数据库的整合

¹ BRIDGES: Biomedical Research Informatics Delivered by Grid Enabled Services [EB/OL]. (2008-8-20), <http://www.brc.dcs.gla.ac.uk/projects/bridges/>

容量的资源的包装器。使用者利用这些包装器可以无缝的使用 **DB2 SQL** 和过程语言进行对外部数据资源的检索，并且可以为这些允许外部访问的数据建立一个联合的视图。

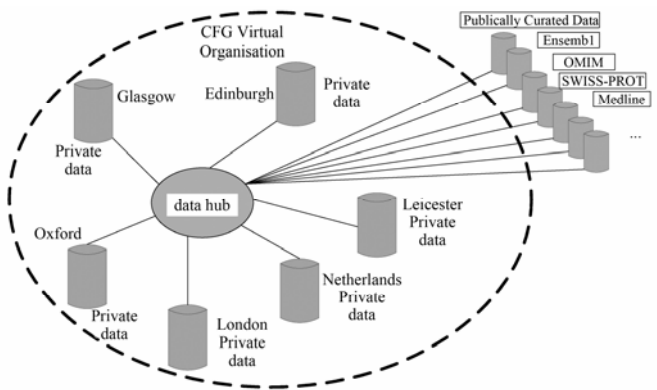


图 7-6 数据集线器（data hub）¹

OGSA 中间件为应用开发者提供了一些服务接口，这些接口用 **java** 开发，允许通过网络进行数据访问和整合。这并不是一个数据库管理系统，它使用一个网格架构来对一套有关的或者无关的数据源进行查询，将结果集和相关元数据最终通过 **SOAP** 返回给用户应用。通过 **OGSA-DAI** 接口（图 7-7），完全不同的各种数据源和资源都可以作为一个单独的逻辑资源来进行处理。

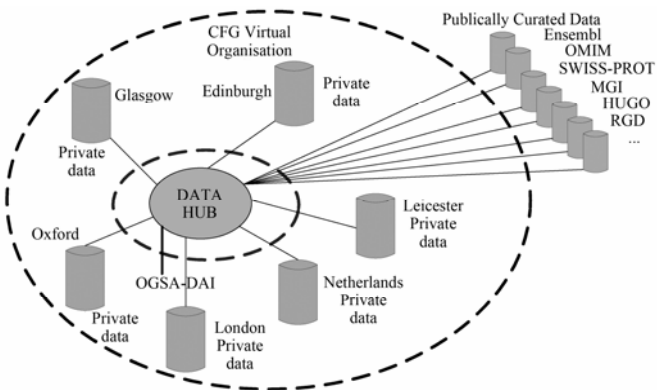


图 7-7 OGSA-DAI的位置²

¹ Richard Sinnott,BRIDGES [EB/OL].(2008-8-20), http://www.jisc.ac.uk/uploaded_documents/bridges-talk-sinnott.ppt

² Richard Sinnott,BRIDGES [EB/OL].(2008-8-20),http://www.jisc.ac.uk/uploaded_documents/bridges-talk-sinnott.ppt

BRIDGES 将 OGSA-DAI 部署到了客户端和信息集成器中间。虽然客户端会因此而略显庞大，但可以使客户端准确地执行网格调用。而且，客户端所发出的请求会发送一个 GridServicesFactory 句柄，后者会将这个请求继续发送到相关的 DB2 存储过程，并且将内嵌的数据放在一个执行文档中返回。这样 BRIDGES 就有了一个可工作的 OGSA-DAI 运行实例，来通过网格服务查询和接收数据，但是实际上，所有的数据集成都是由信息集成器完成的（图 7-8）。

BRIDGES 正在努力开发一种更实际的解决方案，使用 OGSA-DAI 来对 DB2 数据仓库和 MySQL 数据库进行联合查询。

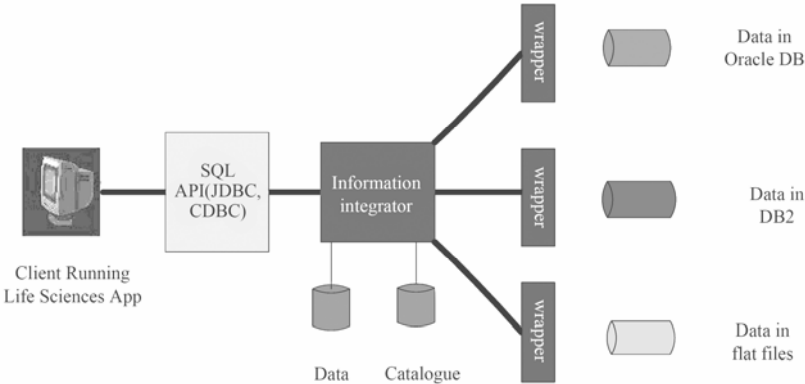


图 7-8 信息集成器¹

7.2.3 BDWorld——大规模数据抽取整合实例

Biodiversity World 项目（BDWorld）正在试图建立一个问题解决环境，目标是为生物多样性研究提供支持，使其能够使用一些网格环境下的通用软件工具来分析从多个数据库和数据存储系统中获得的数据。该项目已经开发了自己的数据访问和通信中间件，以适应由于引入了网格中间件而引起的变化。BDWorld 已经同两个其他项目开展了合作，正在面临着如何使 SPICE 项目中开发的软件访问其他的项目数据库的问题。在 SPICE 项目中，建立了一个 CAS（Common Access System，通用访问系统）网关来对异构的数据库进行打包、访问和连接，形成了一个针对国际 Species 2000 项目的生命目录（Catalogue of Life）原型。在 BDWorld 项目中，这个原型 Catalogue of Life 提供对物种的分类数据，并即将和生物和非生物的数据相关联。这些数据的打包方式与 SPICE（Simulated Pore Interactive Computing Experiment）项目有所不同，这是由于固定的通用数据模式对于 BDWorld 这种大范围使用数据的项目

¹ Richard Sinnott, Grid Services Supporting the Usage of Secure Federated, Distributed Biomedical Data, [EB/OL]. (2008-8-20), <http://www.allhands.org.uk/2004/proceedings/papers/87.pdf>

来说并不合适。

在生物信息学领域，OGSA-DAI 项目的成果，被认为是网格中的分散数据源的访问和整合的重要基石。OGSA-DAI 已经开发出一个软件包，并被 Globus Toolkit 纳入了自己的框架，用来支持访问、检索和处理存储在关系/XML 数据库和普通文件中的数据。

BDWorld 的特点之一是，它使用的是异构的“遗留”下来的数据资源，这些资源也有着不同的结构和数据标准。其中有很多都是 Internet 资源，只能通过 HTTP/XML 协议，甚至是 HTTP/HTML 来进行访问存储。与数据资源异构性相反的是，对这些资源只能进行一些很有限的操作。比如说，为了建立一个研究数据集，需要利用同一个分类号作为检索参数对远端数据集进行重复检索，进而构造出相应的结果集。

BDWorld 也考虑到了这种情况，在一个单独的资源中，任何一种计算任务都可能会被执行。因此这个系统的设计也从争取性能最大化转变到了解决获得资源互操作性的问题上来。BDWorld 建立了一个抽象层 BGI (the BDworld Grid Interface, BDworld 网格接口)，它为所有的 BDWorld 资源提供了一个统一的语法接口，还包括一个启发机制。资源会被包装成符合这个接口的格式，包装好的资源可以通过一个为网格架构特殊设计的适配器同很多个网格或者类似于网格的工具进行交互，其他的 BDWorld 组件都使用这种机制。尤其是 BDWorld 的用户接口是通过 Triana workflow 管理系统提供的，而这个 Triana 管理系统也是一个 BGI 的客户端。

但是获得有关资源的信息，包括数据类型的转换信息，还是很必要的。因此，BDWorld 按照本体建立了一个元数据仓库，目的就是管理资源的不均衡性。有些数据提供者可能会使用非标准的分类名字来对他们的数据进行索引，或者可能使用一个不正规的数据表现方式，因此在对他们提供的数据进行定位和整合的时候，就需要以上这种设计来支持语义等价测试。

为了对远端数据源进行数据访问和获取，BDWorld 资源包装器必须发布这些数据资源的元数据，并且写入 BGI。这个过程包含了对一个统一方法的调用，这个方法可以让 BGI 通过一个统一的方式对所有的资源进行请求。在这个统一的资源请求机制中，BGI 数据调用并不使用标准的 SQL 语言，数据的输入和输出都是以 XML 文档或者一个简单的串的方式进行。

有两种方法可将 OGSA-DAI 引入 BDWorld 项目。第一个就是扩大 BGI，以便将查询纳入工作流中，然后直接被发送到支持 OGSA-DAI 的数据库中，但对于 BDW 协议来说，这将可能涉及很多重大修改，而且还没有考虑到很多这种资源都不是以数据库的形式出现的。另外的一个选择就是在单个包装器范围内提供可以从 OGSA-DAI 获益的基础结构。

在 BDWorld 中，最终他们采用了第二种方式来引入 OGSA-DAI。图 7-9 就展示

了这项研究的基本设计情况。

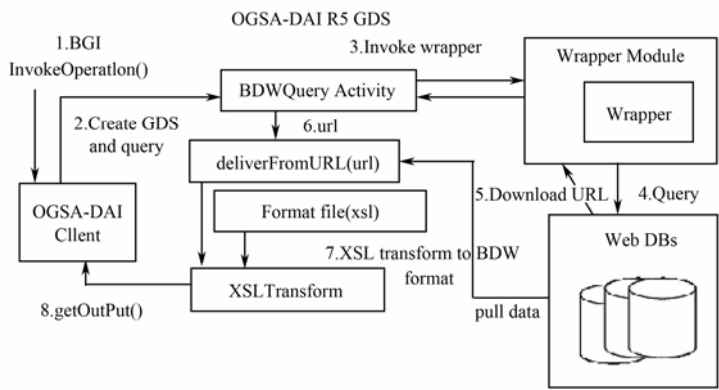


图 7-9 BDWorld¹

这个设计包括了现有的 BGI 设计特点，比如说同样的调用方法、通过资源包装器来访问网络数据库等等。这个设计是建立在一个虚拟的 GDA（Grid Data Service，网格数据服务）之上的，但是这个基础并没有被映射到一个特殊的数据资源上。BGI 可以利用任何一个 GDS 提供的包装器的资源句柄来调用 OGSA-DAI 客户端。接着 OGSA-DAI 客户端建立这个 GDS，并且将行为请求与自己集成到一块来进行操作。BDWorld 查询行为（BDWQueryActivity）调用正确的包装器来对目标数据库进行搜索，返回结果文件的 URL。数据由 URL 获取行为（deliverFromURL Activity）获得，接着输入 XSL 转换行为（XSLTranform Activity）。转换结束之后，输出会被返回给 OGSA-DAI 客户端。图 7-9 演示了如何修改 OGSA-DAI 来满足对网络数据库的访问，但是数据整合过程仍然是由 BDWorld 系统来处理。对于少量数据来说，这个设计是足够的，但是对于大量数据，还必须将这个输出结果输出到缓存中，同时返回一个数据的句柄给 BGI。但是这仍没有利用 OGSA-DAI 的数据整合能力。

图 7-10 表明了如何利用 OGSA-DAI 的灵活的行为框架和如何将一个新的行为加入基础设计来支持数据整合。这个过程需要 BGI 来将一个资源句柄的单子发送给 OGSA-DAI 客户端，并且为发送提供一个目标地点。可以加入一个传统的合并输出行为（MergeOutput Activity）来将分散的网络数据库返回的数据整合起来。为了令带宽最大化，可以利用高性能 GRIDFTP 来将固定的数据集发送给 workflow 单元或者下游操作。

¹ Shirley Crompton etc, Data Integration in Bioinformatics Using OGSA-DAI [EB/OL]. (2008-8-20), <http://www.nesc.ac.uk/talks/ahm2005/500.ppt>

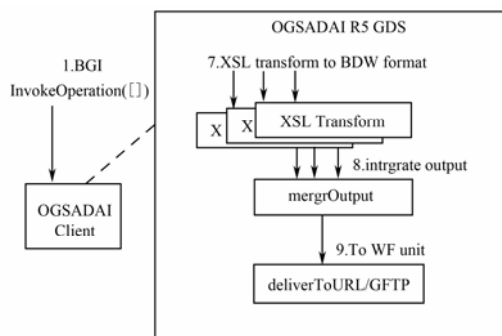


图 7-10 数据整合¹

7.2.4 eSDO——大数据量整合实例

PPARC 资助的 eSDO 项目的目的是利用虚拟观察从太阳动态观测（Solar Dynamic Observatory, SDO）中抽取出数据和算法。AIA（Atmospheric Imaging Assembly, 大气图像集合）和 HMI（Helioseismic and Magnetic Imager, 日震磁性图像记录器）每天要产生几乎 2TB 的原始数据，大量的 SDO 数据就需要特殊的方法来保证有效的数据处理。特别是在 AIA 和 HMI 图片的特征识别、数据的快速摘要生成和数据的快速检索定位之上。

1. SDO Quicklook 格式

每天 SDO 都会产生大量的数据，这就必须找到一个有效的方法来对这些数据存档进行检索。Quicklook，也叫数据浏览，可以让科学家们更有效的识别特殊的数据集。为了 SDO 的其他一些产品，比如说内部轮作和声速图，Quicklook 就会采用相似格式的图像缩略图。大多数的缩略图在斯坦福大学和洛克希德马丁产生。然而，其他的缩略图可以给出一个长时期内太阳的行为的大纲。比如说，HMI 设备每 45 秒可以产生一个次力图，但是英国数据中心每小时才需要一张这样的图片来生成一个本月的太阳活动情况的视频。AIA 数据生成的电影都可以通过将不同渠道得到的 EUV 图片覆盖到相同图片上来进行进一步的压缩。英国太阳协会的调查发现，用户会同时对缩略图和低级 AIA 和 HMI 数据的视频感兴趣。在 ESDO 的第二阶段，有两种存储方式可供选择。第一，缩略图可以通过 AIA 的校准过滤器和 HMI 的 line-of-sight 磁力图以每个小时一张的速度生成。MPEG 电影可以通过每个月的缩略图来生成，然后存储。第二，就是通过所有的 AIA 的校准过滤器和 HMI 磁力图生成缩略图；而且

¹ Shirley Crompton¹, Brian Matthews, Alex Gray, Andrew Jones, Richard White, Data Integration in Bioinformatics Using OGSA-DAI[EB/OL].(2008-8-20),
<http://www.allhands.org.uk/2005/proceedings/papers/500.pdf>

这也不会自动生成一个视频，而是由用户决定这个视频的开始时间、结束时间、所采用的校准过滤器和磁力图的频率，通过一个 CEA 工具来及时生成一个 MPEG 视频。

2. 美国数据中心

最主要的 SDO 数据中心部署在斯坦福大学。在这里安装了 AstroGrid DSA 软件后就可以让英国的网格用户访问 HMI 的问题、AIA 的图片、地图、视频、磁力图和其他的一些高级数据资料。当所有的低端和部分高端数据资料可以在网上获取并且可以在基于网格的磁带仓库获取的时候，一些高级数据资料就会根据用户的需求而及时得到。

通过 AstroGrid 来访问数据资料包括两个软件组件：DSA（Data Set Access，数据集访问）和 CEA（Common Execution Architecture，通用执行架构）。用来描述一个数据资料的原数据和关键词都回在一个数据库中保存，斯坦福大学数据中心会用 Oracle 数据库来保存原数据。用户可以通过利用 web 服务向 DSA 软件提交 ADQL 查询来对数据库进行检索。DSA 将 ADQL 转换成一个 SQL 查询后传递给后端的数据库。一旦元数据结果被数据库返回，DSA 就会将这些表信息转换成 VOTable 格式的 XML 文件。这是这个 VOTable 文件就可以返回给用户，或者作为数据中心的 CEA 工具的输入。这个 CEA 工具就是一个能够将请求数据资源的拷贝分发给用户的在线存储区域的脚本。这个工具可以被设置成使用数据中心的最好的传递协议，比如 FTP，GRIDFTP 等。

所有存储在斯坦福 SDO 数据中心的数据都是可以给公众访问的。大多数的数据资源都会自动生成，存储在一个磁盘缓存中或者磁带中，但是某些高级数据信息就会根据用户的请求及时地生成，并不会在线存储。AstroGrid DSA 软件最多允许对第一层的资源进行访问。第二层的 AIA 资源比如说温度图等会被存储在 Lockheed Martin 的在线磁盘缓存中。这些资源可以通过 VSO（Virtual Solar Observatory）来进行访问。第二层和第三层的 HMI 资源并不是自动生成的，但是也会在用户访问的时候以一种数据库调用的方式建立。斯坦福和 Lockheed Martin 以及它们的共同开发伙伴可以访问这些数据库调用。

3. 英国数据中心

辅助 DSO 数据中心部署在英国，目的就是减少斯坦福的 SDO 数据中心的访问压力，同时减少对横渡大西洋的数据调用的需求。在拉尔（RAL, Rutherford Appleton Laboratory）ATLAS 存储设备已经被定为英国 SDO 数据中心的接班人。这个设备中心可以提供通过磁盘缓存的临时存储和长期在线的磁带存储。预计需要处理三种对象：对斯坦福和洛克希德马丁的 HMI 和 AIA 数据资料的部分镜象、根据特定算法生成的流行数据资料的缓存和缩略资料的存储。如果数据资料在英国数据中心存在的话，这就需要通过斯坦福大学的旁路拉实现对英国用户的快速访问。一个永久的磁

带存储会按照每个小时或者每天这样的频率来对电磁图或者其他图片进行取样，然后将这些样品转换成为视频或者目录，在 SDO 过程中存储起来。

英国数据中心的一个模型会为 SDO 任务期间生成的 HMI 电磁图、多普勒图和强度图提供一个景象。而且英国数据中心会保存两个月内满足以下条件的 AIA 科学数据：1000×1000 的低解析度全局图；500×500 的高解析度活动区域跟踪图；热量图：4~10 张/天；DEM 测量：4~10 张/天；辐照度估计：表状数据；潜在力场磁性外推图：4 张/天；非线性立场磁性外推图，1 张/天。

虽然文件大小也只能当时才能定下来，但是预计基于归档的考虑，三种 HMI 产品——磁力图，多普勒图和强度图——大约每年需要 35TB 的存储空间，而积累下来的 AIA 低端或者高端信息大约每年会需要 55TB 的存储空间。低端 AIA 和 HMI 资料的缩略图只会占用不到 1TB 的空间，但是每个月的按小时取样缩略图生成的 MPEG 集合所占的空间几乎可以忽略不计。

英国数据中心会将描述这些数据资料的原数据存储到一个 MySQL 数据库中，可以通过 ADQL 语言对 AstroGrid DSA 进行检索。CEA 工具可以通过 wget 或者 gridFTP 传输机制在 Atlas 数据存储中心和用户的网格也可以在本地磁盘之间进行传输。英国 DSA 和 CEA 工具可通过 AstroGrid 门户进行访问，而任何可以利用 Web Service 对这些工具进行调用的软件架构也可以访问这些功能。

7.3 数据存储和管理技术

科学数据是科学研究的关键，对如何通过网格环境，对分布在各个实验室中的数据进行有效管理和存储是 e-Science 研究十分关注的一个重要问题。在国外科学数据的管理和存储中，主要有以下几个代表性项目：Geodise database Toolbox（基于网格的工程数据管理方案）；BioSimGRID（基于网格的分布式数据技术）；MySpace（虚拟观测台的数据管理）；Data Portal（门户的方式进行数据管理）；SRB（分布式存储资源的统一代理）。

7.3.1 Geodise：基于网格的工程数据管理

工程设计查询和优化（Engineering Design Search and Optimisation, EDSO）的目标是利用工程建模和分析来实现设计优化。系统设计中的变量需要被系统地进行调整，以提升设计质量。这是一个计算和数据密集型的过程，会在不同的地方产生大量的不同的数据。网格技术的出现提供了一次机会，可以实现这些分布式数据的共享和应用。为了帮助网格之上的这些计算和数据资源，Geodise 项目开发了计算机工具和数据库工具，以帮助工程技术人员构建一个设计环境。Geodise 计算机工具提供了一系列帮助工程人员存取网格计算资源的工具。数据库工具可以帮助用户实现分

布式的数据管理和共享。

一般情况下，一个工程项目产生的数据会被存储在文件系统之中，需要相应的信息对它们进行描述。在数据量非常大的情况下，这些数据将难以查询、共享和重用。这一情况在一个群体希望共同解决一个通用问题的时候变得更加突出。通常情况下，解决的办法是为这些数据添加描述这些数据的元数据，使得这些数据可以通过数据的特性进行定位，而不需要知道它的确切位置。

为了推进元数据在工程环境中的应用，元数据必须简洁且是能够重复嵌入。SRB 提供一个对网络上的异构资源进行统一存取的接口，并且通过 MCAT (Metadata Catalog) 提供了按特性而不是物理位置对数据集存取的功能。但是 MCAT 对于一些特定的应用还存在不足，如对于复杂嵌入数据结构和各种数据类型的支持不足，而这些在工程环境中对一个问题的求解是非常必要的。

为了实现对数据的灵活存取和方便应用，Geodise 开发了数据库工具，它可以让工程人员在 Matlab 环境之下方便地定义元数据，如同定义一个 Matlab 的结构一样。标准的元数据如存档日期、文件大小通过工具包可以产生，可以使工程人员只需要关注定义它的应用系统所需要的特定元数据，并且可以通过授权让一个虚拟组织中的其他人可以存取这些数据。元数据被存储在支持 XML 的关系数据库 (Oracle 9i) 之中，标准的元数据被存储在表中以方便存取，而用户定义的特别元数据被存成 XML 数据以方便灵活定义，并且可以实现 Matlab 结构和 XML 工具包中数据的透明转换。为了实现对数据的查询，Geodise 数据库工具包定义了一种方便的查询模式，可以方便地执行 SQL 和 Xpath 以实现数据查询。为了实现数据的权限管理，Geodise 构建了数据组的概念，并在数据操作上，使用了 Jython 脚本，能够无缝地集成到 Java 代码之中。Geodise 数据库工具包的框架结构如图 7-11 所示。

7.3.2 BioSimGrid: 基于网格的分布式数据技术

生物分子模拟提供了对复杂生物分子系统中能量和动力结构进行模拟的数据信息。这些模拟数据对于推动分子生物学的研究有着重要意义。为了实现将英国的生物分子模拟的共享和集中应用，提出了 BioSimGrid 项目。

BioSimGrid 项目的目标是构建一个通用的数据库实现生物分子的模拟分析，并且能够应用于制药领域。

针对现实中集中式数据的不足 (数据量大、快速存取困难、维护和保存不便)，BioSimGrid 提出了一个基于网格的技术方案，通过构建一个中心元数据库，对各地分布式资源进行有效集成。

BioSimGRID 项目将构建一个英国生物分子模拟的规范数据库，通过分布式的计算机环境，促进研究和合作。在数据库中存在 3 种数据类型的数据 (如图 7-12 所示)。

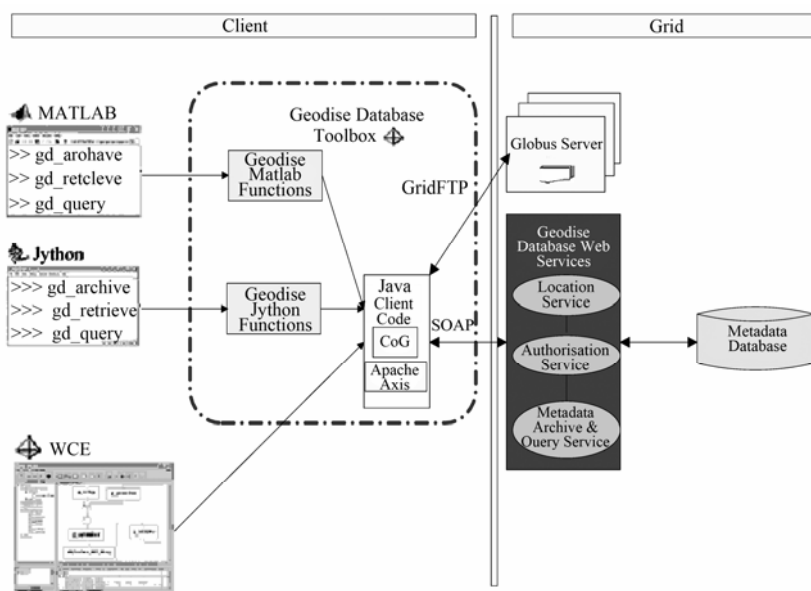


图 7-11 Geodise数据库工具包¹

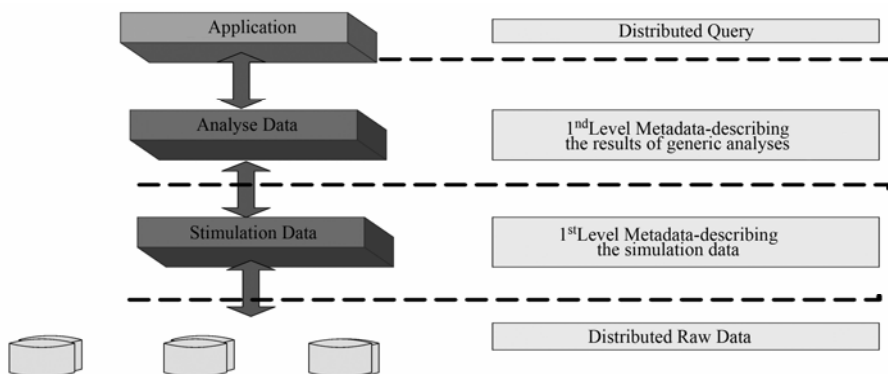


图 7-12 BioSimGRID²

(1) 裸数据：由生物分子模拟产生的数据。

(2) 第一层的元数据：对生物分子模拟数据的裸数据的属性进行总体描述，如

¹ Zhuoan Jiao, Jasmin Wason, Marc Molinari, Steven Johnston & Simon Cox, Integrating Data Management into Engineering Applications, [EB/OL]. (2008-8-20), <http://www.nesc.ac.uk/events/ahm2003/AHMCD/pdf/140.pdf>

² Bing Wu, BIOSIMGRID: A DISTRIBUTED DATABASE FOR BIOMOLECULAR SIMULATIONS, [EB/OL]. (2008-8-20), http://www.biosimgrid.org/public/papers/all_hands_paper_final.pdf

数据所在的地方，模拟器的配置等。

(3) 第二层元数据：对模拟数据的通用分析结果进行描述。它将生成一系列的通用分析工具。

一旦这些数据被组织起来就可以利用工具软件来实现对整个数据库中的模拟数据进行挖掘。图 7-13 是 BioSimGrid 的工作流程。

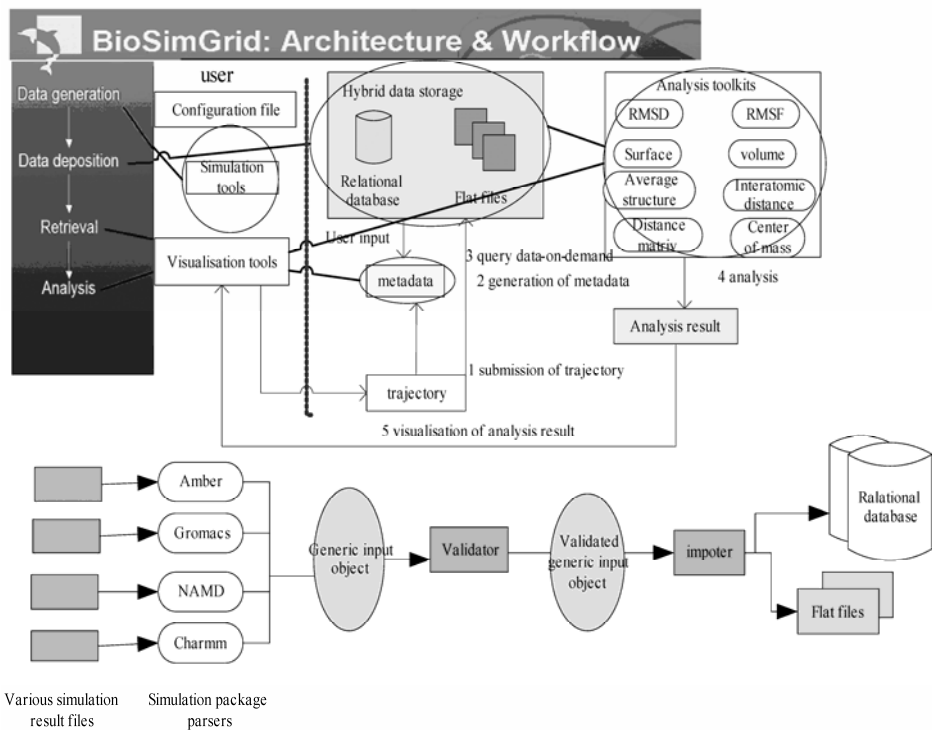


图 7-13 BioSimGrid的工作流程¹

BioSimGrid 基于 OGSi 和 OGSA，实际上通过 Globus 提供一个基于团体的、开放框架和开放源码的服务，促进应用程序处理分布和异构的数据资源。如图 7-14 所示。

从中可以看到，这一系统包括以下组件：

(1) 图形用户界面（GUI）层：是一个基于 HTTP 的 Web 客户机，提供用户与系统的交互。

(2) 服务层：服务层通过基于网格的 Web 服务向生物分子模拟和结构生物研究

¹ Jonathan Essex, BioSimGrid: A distributed database for the storage and analysis of biomolecular computer simulations, 229th American Chemical Society National Meeting, 2005-03-13,[EB/OL]. (2008-8-20), <http://www.biosimgrid.org/public/talks/ACSTalkPresented.ppt>

团体提供数字分析和数据挖掘服务。在服务层，还需要有一些支持服务，如监视、事务处理、分布式查询，在此利用的协议主要有 XML/SOAP。

(3) **Grid 中间件**：这一层包括 GT3 核心的网格中间件，它们提供了核心服务和功能可以实现计算机网络的构建。在 GT3 核心的基础之上，有一些组件实现基本服务，如安全、资源监测、管理、数据存取和通信。**Grid 中间件**中，主要有 **BioSimGrid 数据引擎**和 **SRB**。

(4) **数据库和数据**：数据和数据库是一些分布在不同站点上的资源。网络中间件能够将这些资源的以一种虚拟计算机的方式进行存取。

在这一个多层体系中，相关的应用软件可以被独立的开发，并且作为一个分布式服务集成到系统之中。有利于提高系统的灵活性和可扩展性。

实际上，**BiobSimGrid** 是通过关系数据库（Oracle 10g）来管理数据，以 **SRB** 来实现分布式文件管理，并利用 **MCAT** 来实现元数据管理。如图 7-15 所示。

总体而言，**BioSimGrid** 是一个能够对多种生物分子模拟数据进行分析的工具，它实现了分布式数据库环境，提供了进行数据挖掘、通用数据分析和可视化的相关工具，能够对模拟数据进行标注。它具有存储安全、方便利用的特点，在目前正在牛津，南安普顿，诺丁汉，约克镇和布里斯托得到了应用，新的站点包括伦敦，巴黎，巴塞罗那，里奇兰，圣地亚哥，上海和伊利诺伊。它能够在蛋白质、DNA 等应用领域发挥很大的作用。

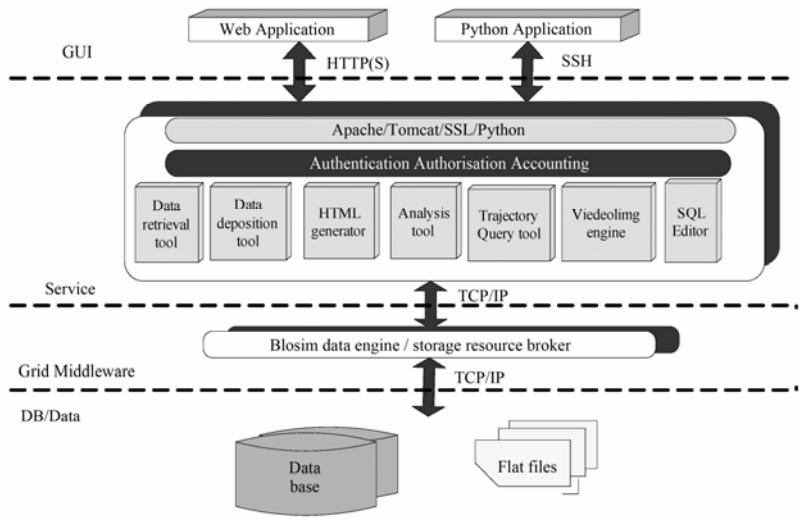


图 7-14 BiobSimGrid的架构¹

¹ Bing Wu etc.,BioSimGrid: Towards A Grid-Enabled Biomolecular Simulations Database, [EB/OL]. (2008-8-20), <http://www.biosimgrid.org/public/posters/AHM2005-poster.pdf>

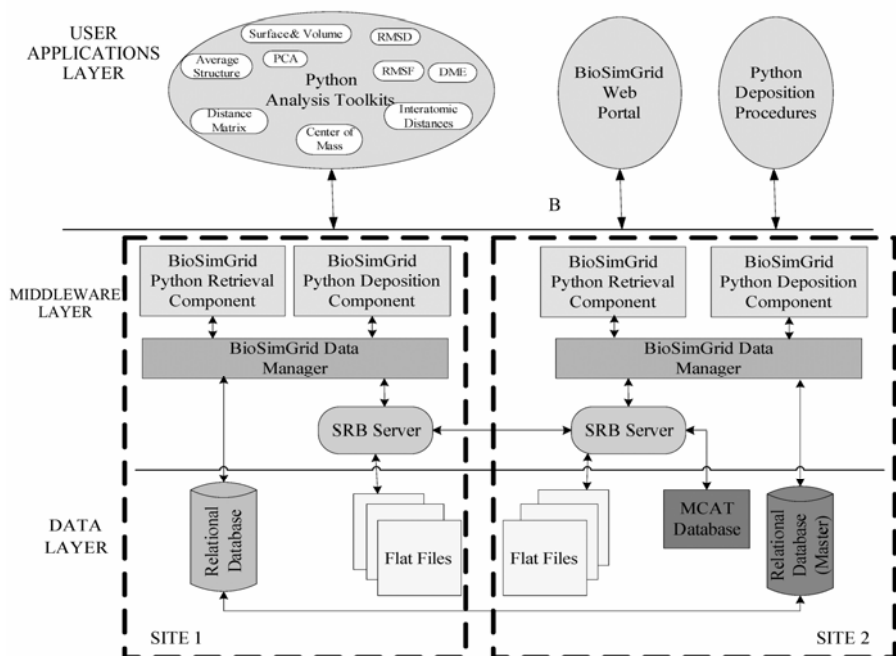


图 7-15 BioSimGrid的文件管理¹

7.3.3 MySpace: 虚拟观测台的数据管理技术

AstroGrid 是英国一个完整的虚拟观测台的开发框架，实现了对存档资源和数据库的无缝存取。许多天文研究需要对档案资源进行查询，从数据库中得到所需的结果。天文学家希望将这些数据下载到它的本地计算机上进行更深入的分析，与其他来源的数据进行集成比较，进行更深入的发掘。在这种情况下，需要一个工作环境，帮助科学家来管理它自己的数据，不论是暂时还是长期存储。MySpace 是 AstroGrid 系统中实现这一功能的一个软件系统。作为一个分布式的 Web 服务协作网络，MySpace 能够让天文学家可以无缝存取地理上分布的工作数据。

MySpace 是 AstroGrid 框架之中的一个重要组件，它完全与正在发展的 IVOA (International Virtual Observatory Alliance, 国际虚拟观测联盟) 接口标准兼容，使得它可以成为一个在多个虚拟观测台中可以即插即用的组件。MySpace 在 AstroGrid 框架中位置如图 7-16 所示。

¹ Jonathan Essex, BioSimGrid: A distributed database for the storage and analysis of biomolecular computer simulations, 229th American Chemical Society National Meeting, 2005-03-13,[EB/OL]. (2008-8-20), <http://www.biosimgrid.org/public/talks/ACSTalkPresented.ppt>

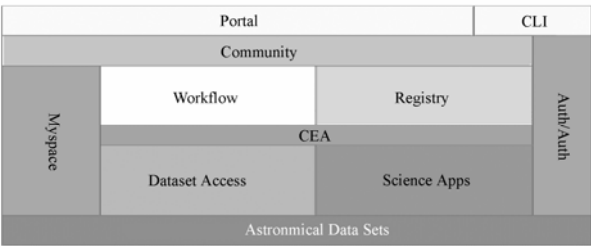


图 7-16 MySpace在AstroGrid框架中位置¹

MySpace 主要有三个功能组件 MySpace Manager，MySpace DataServer 和 MySpace Client（图 7-17 所示）。其中各自功能如下：

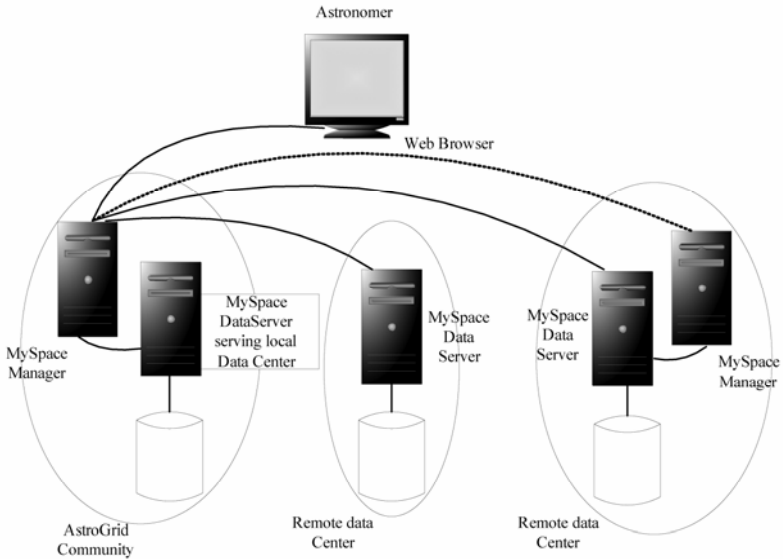


图 7-17 MySpace中的三个功能组件²

1. MySpace Manager。体现了 MySpace 系统的智能，它可以被外部的组件调用，以存取或操纵 MySpace 系统中的文件。它管理和维护了对实际数据集进行描述的元数据，提供了数据集的结构和存取功能。
2. MySpace DataServer。进行数据存储的仓储系统。它被 Manager 调用以进行数据的复制、删除等操作。

¹ The AstroGrid MySpace System,[EB/OL]. (2008-8-20), <http://www.allhands.org.uk/submissions/papers/183.pdf>

² The AstroGrid MySpace System,[EB/OL]. (2008-8-20), <http://www.allhands.org.uk/submissions/papers/183.pdf>

3. MySpace Client。用户和 MySpace 系统的接口，特别是与 MySpace Manager 进行交互，能够让使用者透明地存取 MySpace DataServer。目前情况下，它是 AstroGrid 提供一个特定的 Web 浏览器（称为 MySpace Explorer）。

MySpace 主要提供功能包括：数据存储和共享、安全管理、授权和认证、统计、容错等。

7.3.4 Data Portal: 使用门户进行数据管理的技术

从数据的角度来看，一个科学研究的流程可以被看成如图 7-18 所示的过程。科学研究开始于数据，通过对数据的分析，进而生产出更多的数据。当前的科学家被迫手工地实现对所有实验、数据、计算资源和分析设备的关联，没有太多设备可提供支持。CCLRC 已经发起了一些项目以提供一个对实验、计算资源、数据设备等等的集成方案。实际上，所要 e-Science 的应用系统都需要实现对分布式异构信息/数据的抽取和应用。在 e-Science 环境下，有三个对科学数据的挑战：数据的可存取性、数据的传输和个人数据的管理。

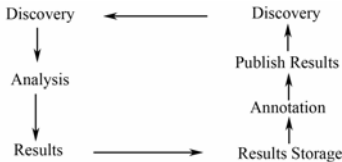


图 7-18 科学研究流程¹

在 CCLRC 中，对数据的管理通过两个基于 Web 服务的门户完成：Data Portal 和 Data Insertion Portal。其中，Data Portal 是对多领域数据的高层存取，它通过元数据描述各种异构的数据信息，并通过 SRB 对物理上分布的数据资源进行管理。Data Portal 与当前的数据目录系统相连接，它的目录中包括了元数据，能够关联原有数据。数据本身可以存储在不同的存资源上，从本地磁盘到数据库，再到各种 TB 级的磁盘系统。Data Insertion Portal 被用于进行数据提交和数据标注。数据标注的结果存储在元数据仓储中，可以让 Data Portal 检索和定位数据。

数据门户的目标是充当用户、设备、数据和其他资源的中间件，它同样是对多种科学数据进行存取的数据网关，希望通过一个单一的接口能够实现对当前科学数据的存取。这一项目的目标是提供一种方便、透明的方式对分布在各个异构站点上的实验、观测、模拟和可视化数据进行存取。同时它还希望连接其他的 Web/Grid 服务，让用户可以更深入地使用所选择的数据，例如进行数据挖掘、模拟和可视化。

¹ Glen Drinkwater, Shoaib Sufi, THE CCLRC DATA PORTAL [EB/OL]. (2008-8-20), [http:// www.allhands.org.uk/submissions/papers/161.pdf](http://www.allhands.org.uk/submissions/papers/161.pdf)

数据门户的结构构架如图 7-19 所示。

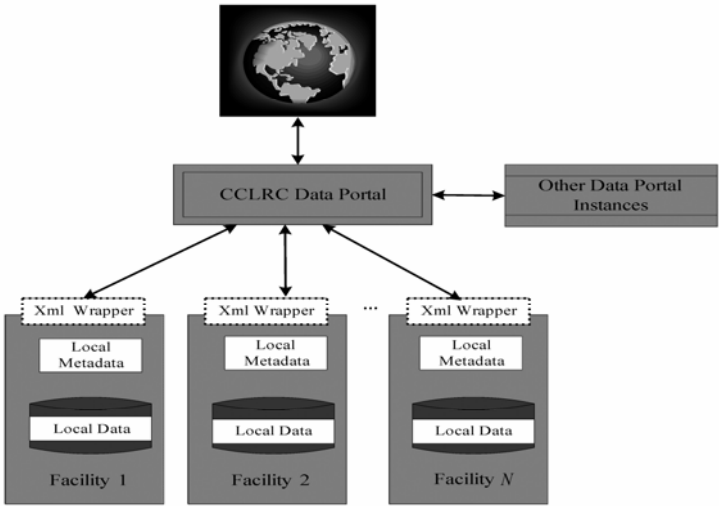


图 7-19 数据门户的结构构架¹

在这一框架中，系统分成三种服务：

（1）核心服务：系统中的每一个设备都提供 Web 服务接口，通过 WSDL 进行描述，SOAP 进行传输，能够数据进行查询和应答。Data Portal 通过 UDDI 服务进行服务的查询和定位。

（2）附加服务：附加的存取控制、认证授权、数据传输、用户管理、设备管理、记账服务。

（3）外围服务：包括 XML 包装器、HPC 门户、可视化、SRB、其他数据门户实例。对于门户需要管理的每种资源，都需要在之前增加一个 XML 包装器，能够将各个设备中的数据包装成了 XML 格式的数据。包装器的框架如图 7-20 所示。

CCLRC 的科学元数据模型（CCLRC Scientific Meta Data Model, CSMDM）是一个面向研究数据集的模型（图 7-21），包括以下几类相关信息：

- （1）主题索引数据：分类法和关键词。
- （2）来源数据：包括这一研究数据集是什么，谁对它在什么时候进行过什么操作等。
- （3）数据收藏信息：关于这些数据的详细说明以及它的布局情况。
- （4）法律信息：关于版权、专利、使用条件的说明。
- （5）相关信息：包括出版、团体信息以及一些与之相关的信息。
- （6）存取条件信息。

¹ Shoaib Sufi, The CCLRC DataPortal, [EB/OL]. (2008-8-20), http://mwi.cdrlr.strath.ac.uk/Documents/CCLRC_dataportal_shoaib_sufi.ppt

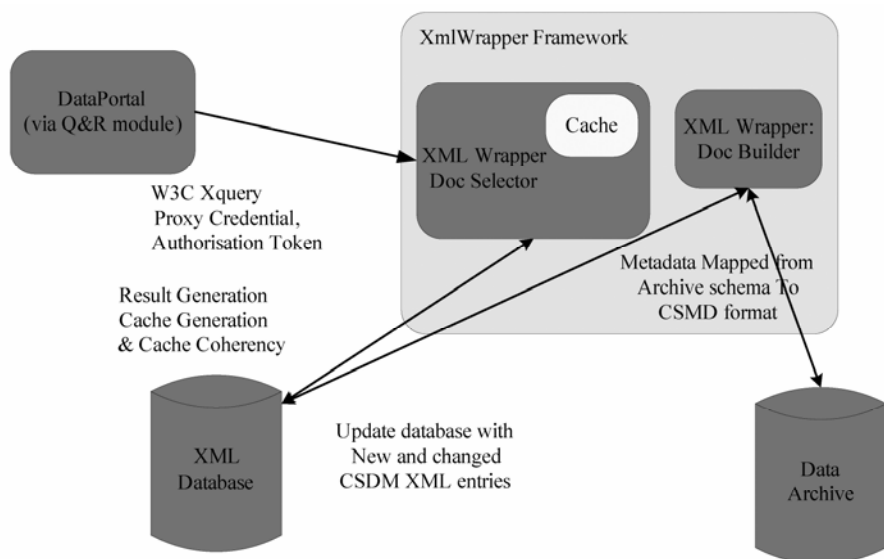


图 7-20 包装器的框架¹

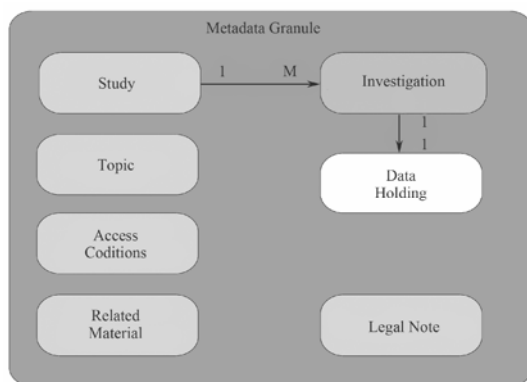


图 7-21 元数据构建图²

在存储方面，Data Portal 利用了 SRB 实现数据的集成，其中每一个资源有一个 SRB 服务，在域 SRB 服务器上，有一个 MCAT（元数据目录），可以将虚拟文件系统中的地址映射到物理的地址之上。如图 7-22 所示。

数据门户需要与 HPC 门户、可视化组件、数据系统及其他数据门户实例进行有

¹ Shoaib Sufi , The CCLRC Data Portal, [EB/OL]. (2008-8-20), <http://www.nesc.ac.uk/events/ahm2004/presentations/161.ppt>

² Shoaib Sufi , The CCLRC Data Portal, [EB/OL]. (2008-8-20), <http://www.nesc.ac.uk/events/ahm2004/presentations/161.ppt>

效集成，形成一个集成的科学研究环境。如图 7-23 所示。

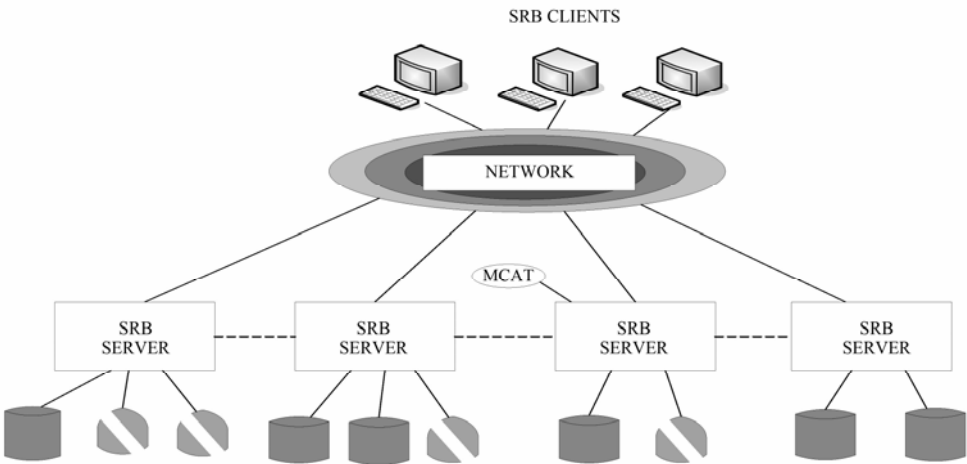


图 7-22 虚拟文件映射¹

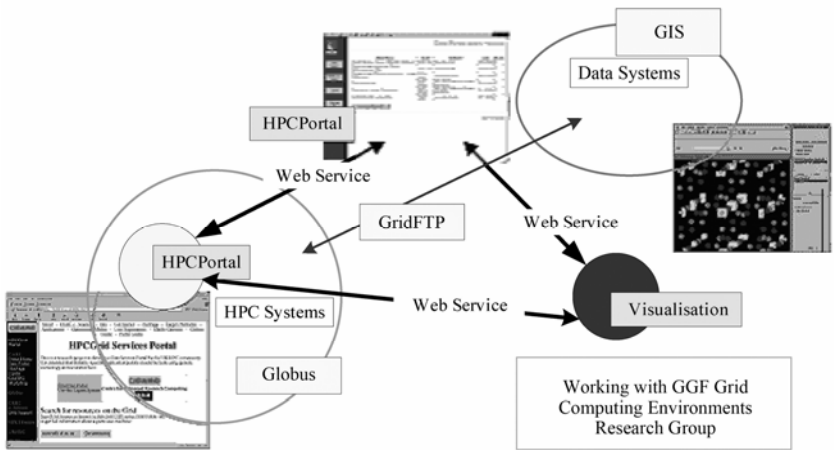


图 7-23 数据门户²

7.3.5 SRB: 存储资源代理技术

数据网格（Data Grid）是跨管理域的高性能计算和信息服务的基础设施，它的

¹ Glen Drinkwater, Shoaib Sufi, THE CCLRC DATA PORTAL [EB/OL]. (2008-8-20), <http://www.allhands.org.uk/submissions/papers/161.pdf>

² Hoaib Sufi. The CCLRC Data Portal, [EB/OL]. (2008-8-20), <http://www.nesc.ac.uk/events/ahm2004/presentations/161.ppt>

目标是将地理上分布、异构的各种高性能计算机、数据服务器、大型检索存储系统和可视化、虚拟现实系统等，通过高速互连网络连接并集成起来，共同支持对数据资源的统一访问，数据网络的体系结构模型如图 7-24 所示。

目前，在数据网络研究领域，比较著名的数据网络系统工具是 Globus 中的数据网络支撑模块和美国圣地亚哥超级计算中心（SDSC）的 SRB 系统。Globus 系统由美国 Argonne 国家实验室和南加州大学联合研制。Globus 对资源的管理、安全、信息服务、数据管理等网络计算的关键技术和方法进行研究，提供了一整套 SDK 和 API，用户可以任意选择其中的工具模块进行高层次的应用开发。Globus 系统最初是面向计算网络的，后来由于数据网络应用的需求迫切，Globus 系统在原有的基础上增加了数据网络的功能，对数据的高速传输、数据复制、数据复制的选择、元数据管理等进行了研究和实现，成为数据网络应用的开发平台。

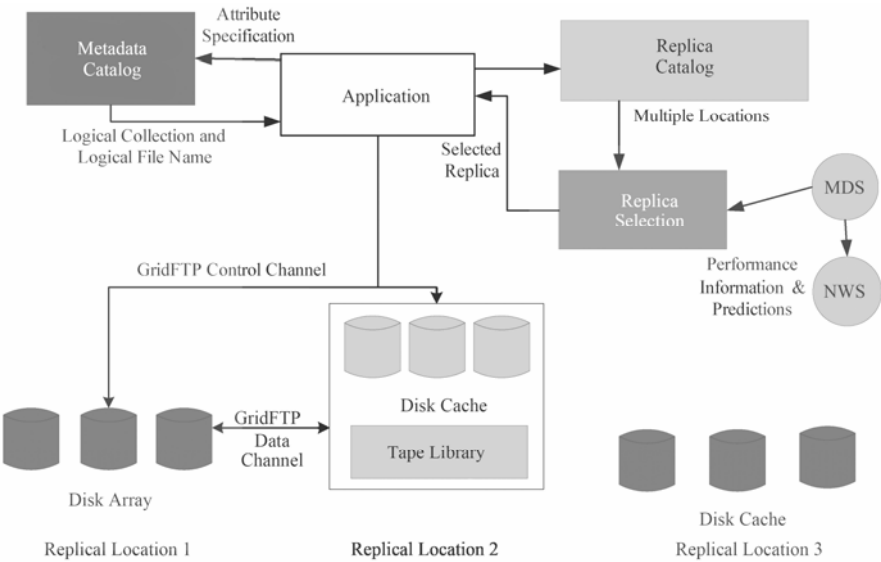


图 7-24 数据网络体系结构¹

存储过程主要使用的网络技术是数据网络，它经由计算网络发展而来，主要是对数据的高速传输、数据复制、数据复制的选择、元数据管理等进行了研究和实现。SRB 存储资源代理系统是用途较广的数据网络软件之一。

1. SRB介绍

SDSC 存储资源代理（Storage Resource Broker, SRB）是一个基于 C/S 模式的中

¹ Hoaib Sufi.The CCLRC Data Portal, [EB/OL]. (2008-8-20), <http://www.nesc.ac.uk/events/ahm2004/presentations/161.ppt>

间件，它为用户提供了一个访问文件系统、档案系统、数据库系统等多种异构存储系统的统一接口，屏蔽了存储系统的异构特性。它支持广域网络环境下多种数据源的访问，提供了复制、复制数据的访问、文件的汇集、分布文件的逻辑集合等功能，提供了一种用于实现数据网格，数字图书馆，以及数据共享、发布和保存的持久稳固存档的抽象机制。目前，SRB 系统已经被五十多个单位使用，包括英国的国家网络、美国的联邦数据库等。

SRB 支持的资源种类可以是 UNIX 下的文件系统、Windows 的文件系统、档案保存系统（如 HPSS、UniTree 和 DMF）、存储在 DBMS（DB2, Oracle, Illustra）中的二进制大对象、DB2 或者 Oracle 中的数据库 SQL 可查询对象，以及磁带库系统，SRB 还包含了一组磁带 I/O 函数以及对 STK 磁带库系统的接口，可以说 SRB 可以起到完整的基础档案存储系统的作用。

作为数据网格软件，SRB 还积极参与网格互联技术发展，比如，早在 1997 年的 SRB1.0 开始就是一个完整的数据网格系统。SRB 可以控制多组分布在本地网络或者广域网网络的客户端和服务端，使其共同协作提供对存储资源、数据和元数据的透明访问。SRB 正在参与很多数据网格研究以及研发、产品合作，其中包括 PPDG、GriPhyN、BaBar、CDL、NASA 等。SRB 在很多项目中或者被作为产品使用，或者被其评估，这些项目来自 NSF、NASA、DOE、DOD、NIH、以及国会图书馆等机构。SRB 支持 GSI（Globus Security Infrastructure）作为一个可选的认证方式，并计划开发兼容 OGSA 的 SRB。

事实上，SRB 既可以像其他网格技术一样被作为中间件来和更高层软件或者其他网格组件来协作，还能够自己独立提供一个完整解决方案，而不需要其他软件（除了需要利用 DBMS 来实现元数据管理）。

2. SRB的性能

SRB 的性能包括它的安全性和吞吐性能。

在安全性方面，SRB 在提供了方便特性和高性能的同时还提供了合理级别的安全保障。其中 Encrypt1 challenge/response 用来防止网络偷听。通常来讲，SRB 和用来存储 MCAT 的 DBMS 以及用来存储数据的物理资源一样安全。用户身份验证和用户主机系统一样安全。并且 SRB 服务器是作为非 root 用户来运行的，所以当操作系统被攻击时，不会暴露操作系统的弱点。从 SRB2.1（2003 年 5 月下旬）开始，SRB 提供了一种加密机制，通过这种机制，SRB 数据文件可以在网络传输和存储过程中被加密。这种机制有效的防止了针对通过 SRB 交换的数据对象的网络偷听，并且加强了存储在各种物理资源的数据对象的安全性。尽管加密和解密过程总是密集计算操作并且会带来不可避免的性能损失，但是这个系统还是高效的。

在吞吐性能方面，传输大文件时，SRB 通常都会比 FTP、SCP 或者 NFS 等快

很多，因为 SRB 具有并行 I/O 能力（网络传输中每个数据流有多个线程）。对于小文件，传输速度会稍微慢一些，因为和 MCAT（尤其是远程 MCAT）的额外交互，但是你可以使用容器或者 Sblast（Bulk load）和 Sblast（Bulk unload）来提高速度。

7.4 元数据管理技术

本节以英国自然环境研究委员会数据网格（Natural Environment Research Council Data Grid。NDG）的元数据的管理为例，分析在 e-Science 环境下元数据管理需要考虑的内容。

1. 科学数据管理的需求

英国自然环境研究委员会（Natural Environment Research Council，NERC）数据网格的目标是提供一个框架以发现和利用自然环境研究委员会所需要的数据。为了支持这一目标，NDG 开发了元数据和数据模型。这一数据模型直接与“利用——呈现”数据有关，它的元数据模型的目标是实现数据的发现和更高层的管理。

这一元数据模型的构建来源于需要对 BADC（British Atmospheric Data Centre，英国大气数据中心）和 BODC（British Oceanographic Data Centre，英国海洋学数据中心）中收集的数据进行管理，但是经过扩展，它可以适应新领域的需要。

NDG 在对科学数据的管理实践中了解到，对科学数据的管理需要满足以下需求：

- ① 在不需要了解数据的存储特性和参数值的情况下，能够发现和存取数据。
- ② 需要对数据进行分领域管理，但是需要超越社区，提供存取功能。
- ③ 可以让不同领域的科学研究获取和使用数据。
- ④ 隐藏数据的异构性，需要能够将来源不同的数据合并为一个统一的结果集。
- ⑤ 可以指定显示处理过程，例如进行子查询，特别是在数据分散在不同数据源的情况下，可实现数据转换和合并。
- ⑥ 可以将特定的数据以特定的格式传递到特定的地方。
- ⑦ 在某些情况下，需要实现服务器端的数据处理。

2. 科学数据和元数据的关系

根据以下要求，NDG 提出了需要将数据和元数据进行区分，通过元数据对数据进行描述和管理的方案。研究人员认为将元数据和数据进行比较是必要的（图 7-25）：

- ① 元数据和数据有明确的功能差异，科学数据的功能是提供科学研究，元数据的功能是发现科学数据；
- ② 将元数据和数据区分开来，可以对元数据进行重用，并将元数据插入到新的应用领域。

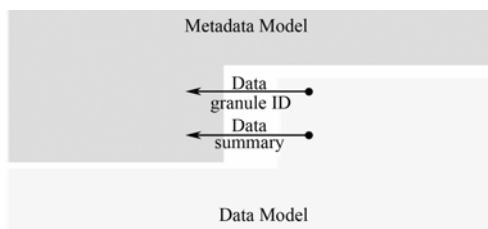


图 7-25 元数据和数据的关系¹

3. 元数据的类型和主要功能

在 NDG 中定义了 A、B、C、D、E 五种类型的元数据（图 7-26）。其作用分别如下。

- A (Archive): 存档格式和利用元数据。与数据一同产生，并且一直与数据一起存在。
- B (Browse): 浏览元数据。关于语义和语法完整描述的元数据。
- C (Comment): 注释元数据。是一种辅助元数据，能够对数据进行标注和说明，一般是在数据被摄入档案之后给出的元数据。
- D (Discovery): 发现元数据。用于发现相关数据集的元数据。通过由数据中心进行管理。
- E (Extra): 其他元数据。针对特定领域的元数据，它可以对核心元数据进行补充。

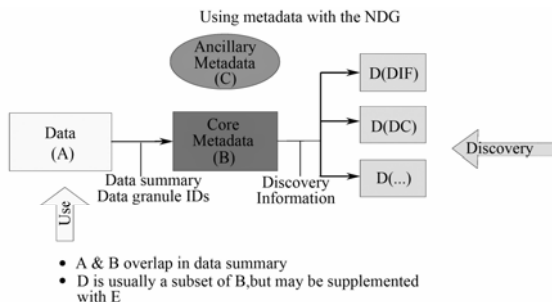


图 7-26 五种元数据及其关系²

4. NDG的元数据概念模型

NDG 提出的元数据的概念模型如图 7-27 所示。在这一模型中，有五种类型的元数据实体，分别有：活动（Activities）、数据产生工具（Data Production Tools）、观测

¹ Kevin O'Neill etc, THE METADATA MODEL OF THE NERC DATAGRID,[EB/OL]. (2008-8-20), <http://www.nesc.ac.uk/events/ahm2003/AHMCD/pdf/129.pdf>

² Kevin O'Neill, Metadata of the NERC DataGrid, [EB/OL]. (2008-8-20), <http://www.nesc.ac.uk/events/ahm2003/presentations/23.ppt>

站（Observation stations）、数据实体（Data entities）和数据集类型（Dataset types）。

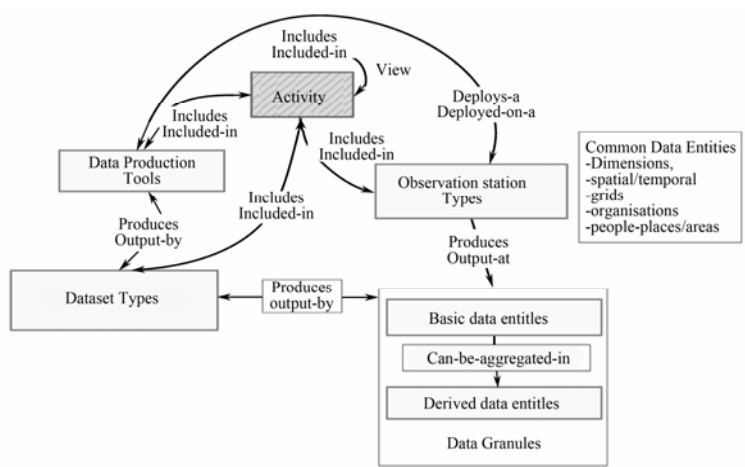


图 7-27 NDG的元数据概念模型¹

5. NDG的元数据管理框架

在地理元数据方面（图 7-28），除了 NDG 元数据之外，还有 ISO 的 211 技术委员会和 191xx 系列标准是关于有地理元数据和服务的标准。在产业界，OpenGIS Consortium（OGC）也在制定可在地理系统中进行互操作的元数据标准。

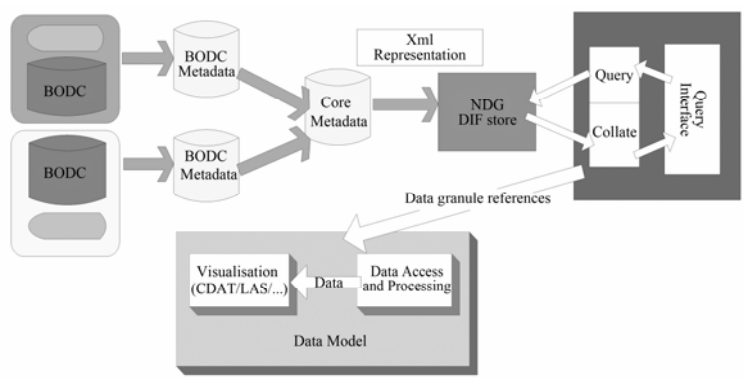


图 7-28 NDG的元数据管理框架²

¹ Bryan Lawrence etc., The NERC DataGrid - Building Bridges for the Environmental Sciences [EB/OL]. (2008-8-20), <http://epubs.cclrc.ac.uk/work-details?w=30086>

² Kevin O'Neill, Metadata of the NERC DataGrid, [EB/OL]. (2008-8-20), <http://www.nesc.ac.uk/events/ahm2003/presentations/23.ppt>

7.5 数据保存技术

英国 JISC 研究支持委员会（The JISC Committee for the Support of Research，JCSR）在研究了目前科学数据的保存情况之后，提出目前最紧急的需要是解决相关的技术问题来保护和发展英国的数字保存。

JCSR 发现英国的科学家都认识到数据革命意味着重要的挑战和机遇。然而，英国还没有对抓住这些机遇做好准备：英国缺少政府层面的支持，不能让研究管理人员参与数据服务以及数据保存框架整体策略的制定；短期保存的资助模式与长期研究的需求相矛盾，目前存在的数据中心通常都是由研究项目的赞助者来支持，对研究人员在数据长期保存方面的工作支持很少；研究人员在工作中通常忽视数据长期保存的困难。

JCSR 认为在未来的十年，按照新的科技设备和实验发展计划，e-Science 涉及的数据储存量将会出现指数级增长。然而针对数据长期保存工作开展得不好、科学数据正处于危险状态的现实情况，JCSR 认为想要从科研投入中充分获利，必须采取进一步的行动，以支持数据的长期保存。

7.5.1 数据保存的三个概念

数据保存是一个比较新的领域，有三个不同的概念。

（1）掌管（Curation）

Curation 是指从数据的被生产出来起就开始的管理和促进其被利用的行为，目标使得数据能够符合现实的需要，或能被用于发现和重用数据。对于动态更新的数据集，需要继续不断丰富它，并且使它适合于当前应用，高级的掌管还包括对数据进行标注和将它与其他出版物相关联。

（2）存档（Archiving）

存档是在 Curation 的基础之上，对数据进行保存的一种活动，它需要确保数据是经过合理选择，得到良好保存的，并且这些数据可以被存取。随着时光的流逝，确保这些数据的逻辑和物理完整性得到维护，具有相应的安全和认证机制。

（3）保存（Preservation）

保存是一种基于存档（Archiving）的活动，数字保存需要解决的问题是即使随着时间的流逝、在技术已经变化的了情况下，还能够对存档的数据进行存取。

尽管这三个概念有所区别，但在实现中，并没有太大的差别。

7.5.2 OAIS参考模型——数字保存的基础

1995 年，在国际标准化组织（ISO）的请求下，空间数字系统咨询委员会（Consultative Committee for Space Data Systems, CCSDS），作为一个致力于合作发

展空间研究中的数据处理标准的国际性空间组织，开始开发一个存档标准以支持数字形式存在的数据的长期保存。在这一行动的基础之上，CCSDS 开发出了一个支持数字存档的参考模型，这一模型创建了描述和说明数据模型的术语和概念，提出了存档框架，指出了在存档环境中重要实体和实体之间的关系，阐述了一个档案系统之中关键的功能和信息组件，最终试图提出一个进行数字信息长期保存的开放框架，能够使现实中的数字档案工作能够建立在这一标准的框架之上。

CCSDS 的这一工作导致 1999 年 5 月开放档案信息系统（Open Archival Information System，OAIS）参考模型的发布。这一参考模型针对数字信息的长期保存和维护的档案系统提供一个概念性的框架。它描述了一个档案系统存在的环境，档案系统的功能组织，以及支持档案处理的信息基础结构。这一参考模型作为 ISO 的推荐草案得到了广泛的评阅，经过扩充，使得它不但适用于空间机构，而且适用于图书馆、档案馆及其他文化传承机构、政府部门和私有机构。新的 OAIS 于 2001 年 7 月发布。

自 OAIS 发布以来，得到了数字图书馆研究人员和组织的广泛关注，很多进行数字资源长期保存系统研究开发的项目，都以 OAIS 作为一个优先遵守的标准规范。本节主要研究 OAIS 具体是什么，OAIS 的目标和范围、OAIS 的环境、OAIS 的主要信息模型和功能模型。

1. OAIS概述

依据 OAIS RM（Open Archival Information System Reference Model，开放档案信息系统参考模型），OAIS 是一个档案系统，它包括一组人员和相关的软硬件系统，这一系统履行对信息进行保存的职责，并且能够使目标团体（Designated Community）获得使用这些信息。这个定义强调了 OAIS 的两个最重要的功能，第一个就是保存信息，例如确保长期保存的安全；第二个就是 OAIS 能够为首要用户或者是目标团体的需要提供对档案信息的访问。

为了与一般的档案系统相区别，OAIS 定义了一个能够被称为 OAIS 的档案系统必须满足的最小功能。它们是：

- ① 这一档案系统必须与信息生产者协商并从信息生产者那里接收合适的信息。
- ② 对所提供的信息的具有充分的控制权限，足够满足进行长期保存的需要。
- ③ 自行确定或与其他相关部门共同确定，哪一个团体是目标团体，从而使存档的信息能够被理解。
- ④ 确保被保存的信息对于目标团体而言是独立可理解的（Independently Understandable）。也就是说，应当确保目标团体在没有信息生产者（专家）的帮助下，仍能够理解信息。

⑤ 遵循书面记载的政策和处理过程，确保信息安全保存，杜绝所有可能控制的意外发生，确保所分发的信息都是经授权分发的信息，它或者是源信息的复本，或

者是其他方式的信息，但通过它可以追溯到源信息。

⑥ 使目标团体能够利用到被保存的信息。

总而言之，OAIS 这一术语，或者说是 OAIS 中提出到的档案系统，意味着这一个档案系统这一档案系统致力于保存数字信息，并且在长期的一段时间之后这些信息仍然可用，同时这一档案系统还履行了上面所说的六个强制性职责。

2. OAIS参考模型的目标和范围

OAIS 参考模型同时支持数字化的和物理存在的存档信息。它特别对数字化信息给予了较多关注。这些数字信息或者是作为档案的主体，或者是作为对数字化资源以及物理存档资源的支持信息存在。

这一参考模型的目标在于：

① 提供一个框架以了解和认识到与数字信息的长期保存和存取访问相关的一系列存档概念和思想。

② 为非档案机构提供一系列的概念和方法体系，以有效地推动这些机构参与数字保存过程。

③ 提供了一个框架，包括术语和概念体系，以描述和比较当前和未来存档活动的框架和运作方式。

④ 提供了一个框架，以描述和比较不同的长期保存策略和技术。

⑤ 为档案中保存的数字信息的数据模型进行比较的提供一个基础平台，并且讨论了随着时光的流逝，数据模型和相应的信息如何演化变迁。

⑥ 提供了一个可扩展的基础框架，使人们可以在将来在其基础之上对非数字形式的信息（如物理媒体或物理样品）的长期保存进行扩充。

⑦ 对数字信息长期保存和存取的组成因素和过程进行详细论述，制造舆论环境，促进市场的扩大，得到更多厂商的支持。

⑧ 提供一个标准指南，让人们据此判断和识别符合 OAIS 规范的系统和产品。

在 OAIS 参考模型中，提供了一个完整的档案信息保存功能，它包括摄入、档案存储、数据管理、存取和分发。它同时论述了数字信息从一种媒体或格式到另一媒体或格式的移植以及进行信息表示的数据模型，信息保存中软件的作用，档案系统之间数字信息的交换等。它还确立了各个档案功能内部和外部的接口，以及一系列在这些接口之上的高层服务。

OAIS 的蓝皮书“Reference Model for an Open Archival Information System (OAIS)”一共分 6 章，除第一章为引言之外，其他各章分别阐述了 OAIS 的概念、OAIS 的职责、OAIS 的详细模型（OAIS 的功能模型、OAIS 的信息模型、OAIS 中信息包的转换）、OAIS 的保存思路（现实中如何进行数字信息和能够访问这些数字信息的存取服务的保存）、OAIS 开放存档系统之间的互操作性。下面将对 OAIS 的环境组成和 OAIS 的参考模型进行分析。

3. OAIS的环境

OAIS 认为，一个 OAIS 是一个置身于生产者、消费者和管理者之间的一个存档体系。如图 7-29 所示。

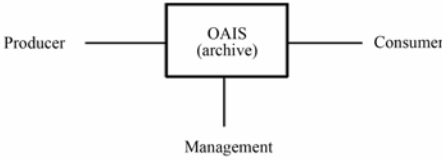


图 7-29 OAIS 的环境模型¹

围绕着 OAIS 的，进行档案存储、管理和利用的角色分别有生产者、管理者和消费者。其中各自作用如下：

- ① 生产者是提供需要保存信息的人或者是客户系统。生产者通过 OAIS 的摄取过程 (ingest) 向 OAIS 提交要保存的信息，包括元数据。这个过程接受提交的数据。
- ② 档案系统中存储。生产者与 OAIS 的相互作用通常被描述为提交协议 (submission agreement)，具体包括提交的信息类型，期望生产者能提供的元数据，从生产者到存档的实际传送管理。
- ③ 管理者由设置 OAIS 整体政策的人组成。这一角色确立档案收集的范围，明确档案条目的保存许可，指明资金来源以及监管使用的情况，定期评论 OAIS 的政策和运作。

④ OAIS 的消费者指的是与 OAIS 服务交互以获取特定保存信息的人或系统。在 OAIS 消费者中，包括了被称为“目标团体”的特殊消费者。

管理者、生产者、用户以及 DC 在 OAIS 中更多的代表了功能性，而不是单纯的组织角色。所有这些角色可以被包含在单一的组织架构内，也可以分布在多个不同的组织中。

4. OAIS 的参考模型框架

在 OAIS 的参考模型框架中，OAIS 的制定者提供了一系列的术语和概念，来描述 OAIS 的功能实体和 OAIS 中存在的信息。

图 7-30 描述了 OAIS 参考模型的框架结构。在 OAIS 中定义了六个功能实体和 OAIS 中的三种信息包。

OAIS 中 6 个功能实体分别为：

¹ Consultative Committee for Space Data Systems (CCSDS), Reference Model for an Open Archival Information System (OAIS)[EB/OL]. (2008-8-20), <http://public.ccsds.org/publications/archive/650x0b1.pdf>

① 摄入功能。这一实体提供的服务和功能是从信息生产者那里接收提交信息包（Submission Information Package, SIP），并且对内容进行准备，以便所提交信息能够在档案系统中进行存储和管理。Ingest 的功能包括：接收 SIP 包、对子 SIP 包进行质量确认、生成符合档案系统数据格式和文件标准的存档信息包（Archival Information Package, AIP）、从 AIP 中抽取出描述信息（Descriptive Information）以存入档案数据库，并且协调对档案存储和数据管理的修改工作。

② 档案存储（Archival Storage）功能。这一实体提供的服务和功能是存储、维护和检索 AIP。档案存储的功能包括从 Ingest 接收 AIP，并将它们存储到永久存储系统之中、管理存储系统的组织结构、对存储有档案的媒体进行翻新、执行日常的维护工作和特殊的错误检查、提供灾难恢复能力、并且为 Access 提供 AIP 以实现档案提取。

③ 数据管理（Data Management）功能。这一实体提供的服务和功能是植入、维护和存取那些标识并记录档案馆藏的描述信息（Descriptive Information）以及对档案系统进行管理的管理数据。数据管理的功能包括管理档案数据库（维护数据库中概念模型及视图的定义，维护系统的参照完整性等）、执行数据库更新（装载新的描述信息或档案管理数据）、对管理数据提供查询功能以产生结果集，并从这些检索结果集中生成查询报告。

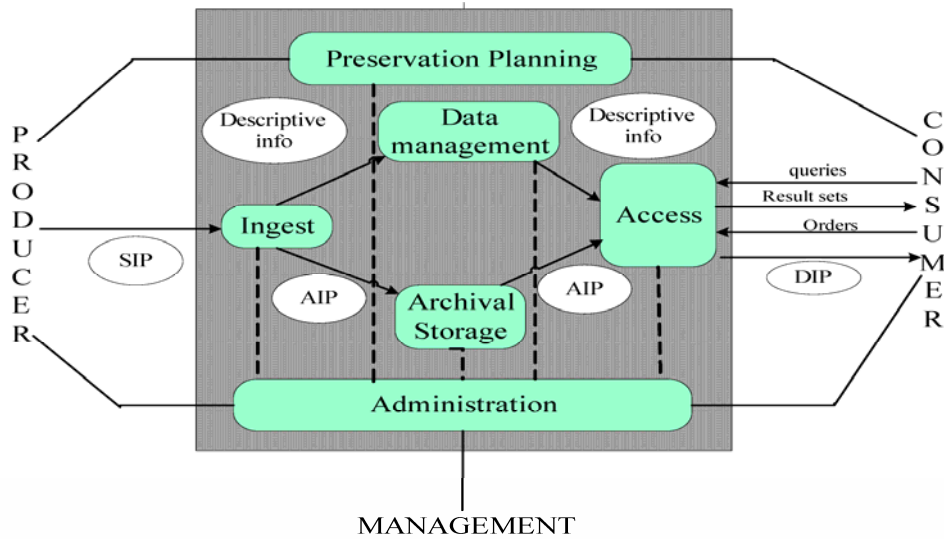


图 7-30 OAIS的参考模型框架¹

④ 管理（Administration）功能。这一实体提供的服务和功能是对整体的档案系

¹ Consultative Committee for Space Data Systems (CCSDS), Reference Model for an Open Archival Information System (OAIS)[EB/OL]. (2008-8-20), <http://public.ccsds.org/publications/archive/650x0b1.pdf>

统提供操纵管理。管理功能包括与信息生产者进行商讨以决定提交协议，对提交进行审计以确保提交内容符合档案标准、对系统的软件环境进行管配置和管理。它同样还包括系统工程学的一些功能，如监测并提高档案系统运作、详细记录、报告档案的内容，并且对档案的内容进行移植和修改。它也包括建立并维护档案标准和政策、提供客户支持、激发存储请求；

⑤ 保存规划（**Preservation Planning**）功能。这一实体提供的服务和功能是监测 OAIS 的环境，提供相关建议以确保在 OAIS 中存储的信息在长期一段时间之后，仍然能够被相应的目标用户所访问，即便是在原有计算环境都退化了之后。保存计划包括的主要功能有对档案系统中的存储内容进行评估，并且周期地提供存档信息建议以移植当前的档案馆藏；对档案系统的标准和政策进行建议；监测技术环境和目标用户的服务需求及知识背景的变化。保存计划同样包括设计信息包（**Information Package, IP**）的模板，提供设计帮助和相关评价以对这些模板进行专业化处理形成特定的 SIP 和 AIP。保存计划还包括开发详细的移植计划，开发移植软件原型系统和相应测试计划以实现管理移植目标；

⑥ 存取（**Access**）功能。这一实体提供的服务和功能是支持消费者，使消费者可以判定、了解、定位以及存取存储在 OAIS 中的信息，让消费者可以请求并接收信息产品。存取功能包括与消费者进行通讯以接收请求，对于某些特别保护的信息进行管理控制以限制其访问，协调请求的执行以实现请求的完全响应，生成响应（分发信息包 DIP，结果集，报告）并且将这些响应传送给消费者。

5. OAIS信息包及其描述

除了对 OAIS 中的功能实体进行定义之外,OAIS 的制定者还通过信息模型，对 OAIS 中存在的信息进行了描述。

在 OAIS 中，信息对象（**Information Object**）由数字对象（**Data object**）本身和它的描述信息（**Representation Information**）组成。也就是说“数据在经其描述信息的解释之后得到有用的信息”。为了成功地保存信息对象，需要让 OAIS 明确地标识和理解数字对象及其相关的呈现信息。描述信息又可以分为结构描述信息和语义描述信息两种。

OAIS 提出了信息包的概念来说明档案生产者向 OAIS 的提交过程和 OAIS 向档案消费者的分发过程。一个信息包是一个包含名为内容信息（**Content Information**）和保存描述信息（**Preservation Description Information, PDI**）两种信息对象的容器。内容信息和保存描述信息被认为是通过打包信息（**Packaging Information, PI**）进行封装和标识的，而为了使打好的信息包可以被了解和发现，需要通过描述信息（**Descriptive Information**）对信息包进行描述。

OAIS 对提交给它的信息包、它所存储的信息包、以及它分发给其他消费的信息包进行了区分，分别将这些信息包叫做提交信息包（**Submission Information Package,**

SIP)、存档信息包 (Archival Information Package, AIP) 和分发信息包 (Dissemination Information Package , DIP)。三种包的作用如下:

① 提交信息包 (SIP) 是信息生产者提供给 OAIS 的信息包。它的格式和具体内容通常情况下是生产者和 OAIS 系统之间协商的结果。多数的 SIP 通常有一些内容信息和 PDI, 但它可能需要多个 SIP 来形成一个完整的内容信息和相应的 PDI, 从而进一步形成一个 AIP。一个 SIP 也可能包括需要被包括在多个 AIP 中的信息。SIP 对应的打包信息通常会以某种形式存在。

② 在 OAIS 中, 一个或多个 SIP 需要被转换成为一个或多个 AIP 以进行保存。AIP 有一系列完整的 PDI 和相关的内容信息。一个 AIP 还可能会由多个其他的 AIP 组成。AIP 的打包信息 (PI) 需要遵守 OAIS 内部的标准, 并且在 OAIS 的管理下, 它还可能会发生变化。

③ 根据消费者的请求, OAIS 需要以分发信息包 (DIP) 的方式提供一个 AIP 的所有或者部分内容给消费者。一个 DIP 也可能包括多个 AIP, 并且它可能有、也可能没有完整的 PDI。打包信息 (PI) 需要以某种方式显性出现, 以使消费者可以明确地辨别出所要的信息。根据分发媒体和消费者需求的不同, 打包信息可能以多种方式存在。

在 CCSDS 和 ISO 的推动下, OAIS 参考模型成为了数字信息长期保存系统普遍遵从的标准规范, 来自图书馆、档案馆、科学数据、文化遗产等领域的长期保存项目、系统都纷纷以其为准, 进行系统的规划和设计。

7.5.3 DCC数字保存的三个阶段

数据是一种输出。输出这些数据的过程通常需要很大的投入, 产生和收集数据需要花费很多的经费, 并且有些数据是唯一的, 无法重复产生。因此, 科学数据是值得进行长期保存的, 以 NERC (Natural Environment Research Council, 英国自然环境研究委员会) 为例, 它在数个世纪之前就已经开始了科学数据的收集和保存工作。

图 7-31 以英国生物技术与生物科学研究委员会为例 (Biotechnology and Biological Sciences Research Council, BBSRC) 来说明 NERC 的资助模式, 它包括输入、处理和输出三个阶段。

在输入阶段, 研究机构、用户团体、政府机构和政策制定者提出他们的研究计划, 选择优先发展领域, 并提交英国生物技术与生物科学研究委员会。

在处理阶段, 英国生物技术与生物科学研究委员会组织同行评议, 从中选择最终的资助项目进行资助。

在输出阶段, 英国生物技术与生物科学研究委员会对研究结果进行传播利用, 包括: 促进对未知领域的理解, 应用于新产品的开发, 实现新工艺的应用, 进行技术改造提高利润, 改进的健康, 提高生活质量, 培养人才等。研究结果的传播和利

用情况还会进一步被反馈给输入和处理阶段。

在所有的输入、处理和输出三个阶段，英国生物技术与生物科学研究委员会都会保存相关的科研文档。

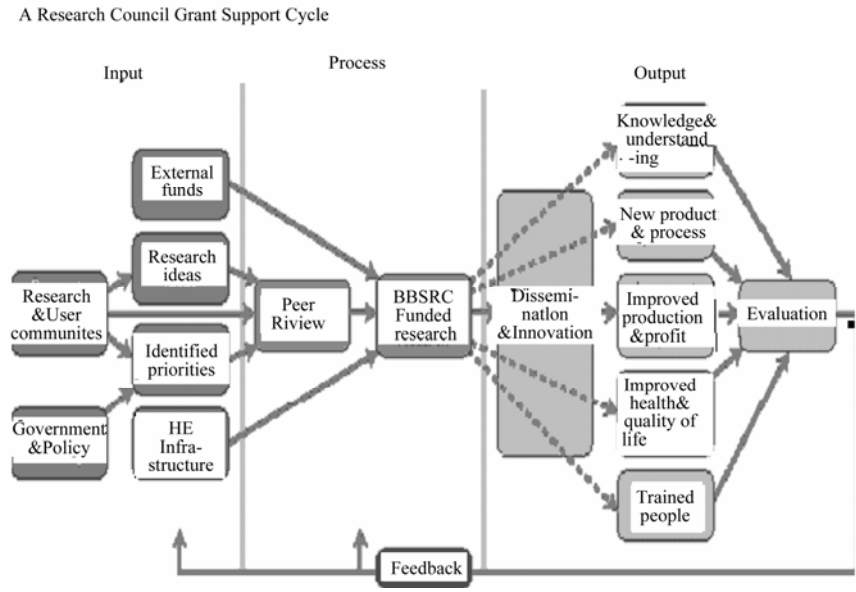


图 7-31 研究委员会保存周期的输入输出

2004 年英国 JISC 和 e-Science core 计划发起了数据掌管项目，构建了三个数字掌管中心，分别为位于爱丁堡大学（与格拉斯哥大学共建 Digital Curation Center）、巴斯（UK Office for Library Networking, UKOLN）和研究委员会中心实验室委员会（Council for the Central Laboratory of the Research Councils, CCLRC）。

Digital Curation Center (DCC) 的主要目标是完整收集和保存科学研究成果。DCC 以一个当前研究和成果出版发布的过程模型为基础，在此之上提出了增加数据存档过程和完整数据保存活动过程来对其进行扩展的想法。

图 7-32 展示了传统研究过程的模型，揭示了信息的转移和提炼过程，从科学家的原始数据获取，到数据分析和研究，形成第二层次的结果数据。之后，研究成果被提炼成为出版物，通过预印本机制和同行评议机制，进行学术出版。出版之后的期刊论文被索引、文摘等二次文献库收录，二次文献库又进一步为综述等三次文献提供了素材。所有的形成的相关文献都被存档。而来自出版循环过程的信息又回到科学家以及科学研究之中。同时，这些资料也被传播给更广泛的社会大众，比如通过图书馆、公众、行业协会和国际同行等。值得注意的是，为了满足合法保存或者专业学术需要，例如，为了保障研究管理的有效性，即使是在这个阶段的主要研究数据，也可能需要保存。

DCC 称这个最简单的过程为第一阶段的保存。

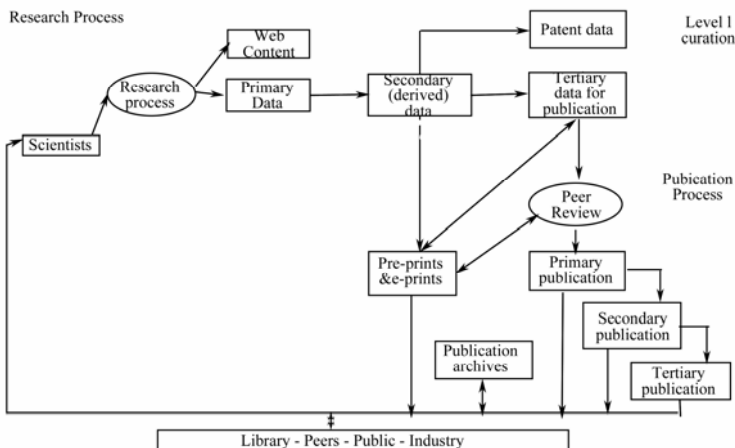


图 7-32 传统学术信息流——第一阶段的保存¹

在某些领域，已经开始了一种转变趋势，即将一次和二次数据保存在专业存档机构中。DCC 在前一个模型的基础上增加了一个数据存档过程。图 7-33 展示了引入了额外的数据流和过程之后的模型，它也展示了新模型和传统研究模型的在数据处理能力上的不同，例如，新模型可以使用数据来做出新的发现。DCC 称其为第二阶段的保存。

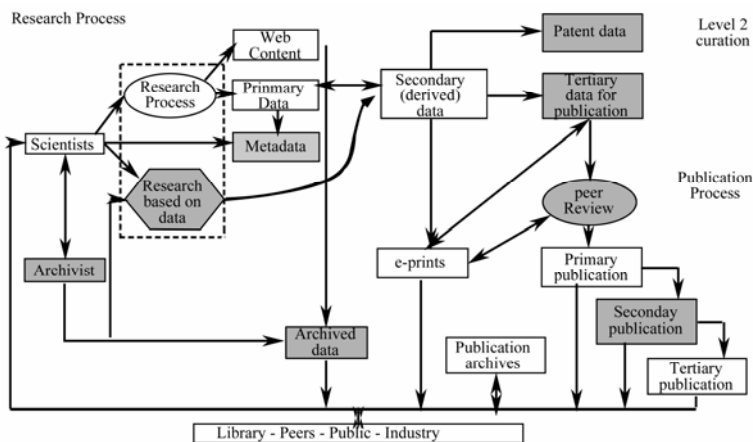


图 7-33 数据存档中的信息流——第二阶段保存²

¹ Data curation for e-Science in the UK: an audit to establish requirements for future curation and provision,[EB/OL]. (2008-8-20), http://www.jisc.ac.uk/uploaded_documents/e-ScienceReportFinal.pdf

² Data curation for e-Science in the UK: an audit to establish requirements for future curation and provision,[EB/OL]. (2008-8-20), http://www.jisc.ac.uk/uploaded_documents/e-ScienceReportFinal.pdf

在图 7-34 中出现了一个新的角色，就是数据掌管者。总的来说，这个角色需要和数据生成者一起来准备对科研数据进行存档（例如生成元数据来确保数据可以被找到，并且在未来能够被复原和应用）。数据掌管者的出现表示着在他们的管理之下，可实现可持续的信息长期保存和重用，DCC 称其为第三阶段的保存。

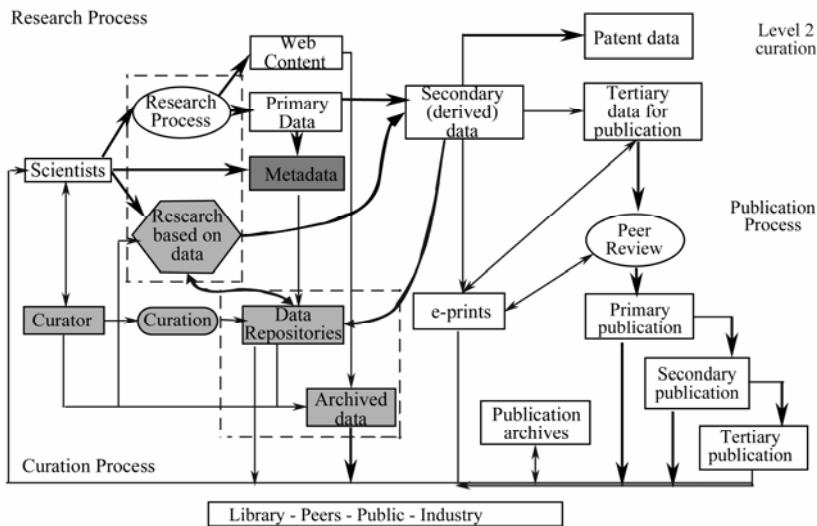


图 7-34 数据 curation 的信息流——第三阶段保存

7.6 数据分析处理技术

7.6.1 两种数据分析技术

数据收集与维护的最终目的是为了供人们使用，科学研究除了获得数据的表层信息之外，更重要的是获得数据之间的内在关系和隐含信息。这需要从大量数据资源中挖掘有意义的知识。

从数据的分析处理方面，目前主要存在着两条研究路线。一条路线基于数据库中的知识发现（Knowledge Discovery in Database）和数据挖掘（Data Mining），主要研究的是从结构化的数据中发现新的知识；而另一条路线基于自然语言处理（NLP）和文本挖掘（Text Mining），主要研究如何从非结构化或半结构化的数据（如 Word、HTML、或 PDF 文件）中发现新知识。

1. 基于数据库中的知识发现和数据挖掘的数据分析技术

KDD（Knowledge Discovery in Database），即数据库中的知识发现。也有的人直接将 KDD 等同于数据挖掘（Data Mining）。数据挖掘是从大量的、不完全的、有噪

声的、模糊的、随机的数据集中识别有效的、新颖的、潜在有用的，以及最终可理解的模式的非平凡过程。

1989 年 8 月，在美国底特律的第 11 届国际人工智能联合会议的专题讨论会上召开了第一届 KDD Workshop，1991、1993、1994 年又接着举行 KDD 专题讨论会。从 1995 年开始，每年都举办一次 KDD 国际会议。从 1997 年开始，KDD 也拥有自己的专门杂志《知识发现与数据挖掘》（《Knowledge Discovery and Data Mining》）。1995 年以来，国外在知识发现和数据挖掘方面的论文非常多，已形成了热门研究方向。

基于 KDD 的数据分析是从大量结构化数据中提取出可信的、新颖的、有效的并能被人最终理解的模式的高级处理过程，通过综合运用统计学、模糊数学、神经网络、机器学习和专家系统等方法，从大量的数据中提炼抽象的知识，揭示出蕴涵在数据背后的客观世界中的内在联系和本质规律，实现知识的自动获取。

基于 KDD 的数据分析过程是多个步骤相互连接、反复进行人机交互的过程。它的基本流程如下：

- ① 问题定义。了解相关领域的有关情况，熟悉背景知识、明晰用户要求。
- ② 数据提取。根据要求从数据库中提取相关的数据。
- ③ 数据预处理。主要对前一阶段产生的数据进行再加工，检查数据的完整性及数据的一致性，对其中的噪声数据进行处理，对丢失的数据进行填补。
- ④ 数据挖掘。运用选定的知识发现算法。从数据中提取用户所需的知识，这些知识可以用一种待定的方式表示或使用一些常用的表示方式。
- ⑤ 知识评估。将发现的知识以用户能了解的方式呈现，根据需要对知识发现过程中的某些处理阶段进行优化，直到满足要求。如图 7-35 所示。

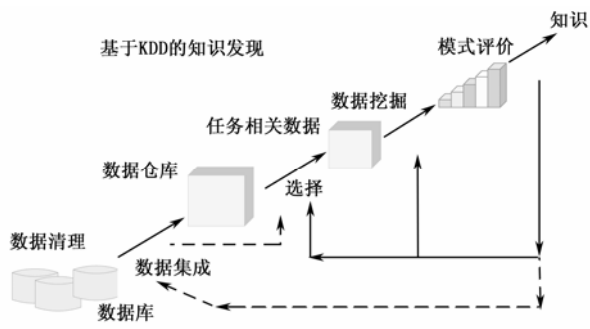


图 7-35 基于 KDD 的知识发现

2. 基于自然语言处理的数据分析技术

与 KDD 主要针对数据库中有结构的数据进行知识发现不同，文本挖掘利用自然语言处理技术，针对更为广泛存在的无结构或半结构的数据（如 HTML 文件或 PDF 文件）进行知识发现。最初的文本挖掘希望“从大量的非结构化的数据中标识并抽

取出事件的趋势和模型，并它们转换成为有用并可理解的信息”。

信息抽取（Information Extraction）是目前日渐成熟，并得到越来越多人关注的文本挖掘方式。经过 MUC（Message Understanding for Comprehension），MET（Multilingual Entity Task Evaluation），ACE（Automatic Content Extraction）等研究测评活动的推动，信息抽取现已发展成为了包括以命名实体识别（Named Entity Recognition）、多语种实体识别（Multi-lingual Entity Task）、模板元素（Template Element）信息抽取、参照（Coreference）、模板关系（Template Relation）信息抽取、情节模板（Scenario Template）信息抽取等不同层次信息内容抽取为对象的完整的研究领域。

近几年来，随着自然语言处理（Natural Language Processing, NLP）、人类语言技术（Human language technology, HLT）、计算机语言学（Computational Linguistics, CL）、知识工程（Knowledge Engineering, KE）、知识管理（Knowledge Management, KM）、语义网络（SemanticWeb）、智能代理（Agent Based Computing）、Web 智能（WebIntelligence）的发展。欧洲美发达国家明确提出了“知识技术”这一概念。

知识技术针对目前的信息过载和数据泛滥的情况而提出，在信息技术的基础之上，引入了智能技术，其主要目标在于创建新一代的、基于语义的并且能够支持上下文关系理解（semantic-based and context-aware）的知识系统，通过它能够自动地进行知识的获取和建模、表示和可视化、解析和共享。

知识获取和知识建模是与知识发现密切相关的两个研究内容（图 7-36）。

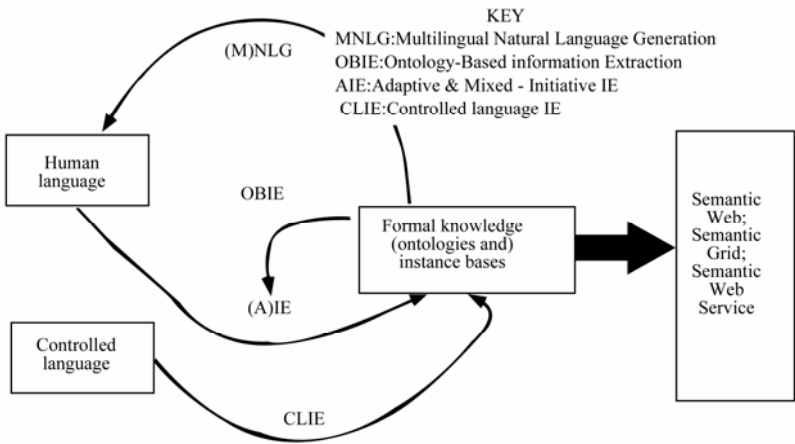


图 7-36 基于信息抽取的知识获取和知识建模¹

¹ Hamish Cunningham, Kalina Bontcheva, Diana Maynard, Valentin Tablan, Human Language Technology for the Semantic Web, [EB/OL]. (2008-8-20), <http://www.gate.ac.uk/sale/talks/esws2004-tutorial.ppt>

知识技术所提的知识获取的目标在于从大量的可获得的信息中,提取出相关的知识并将其转换为可用的方式,特别是研究如何从非结构化的、多种类型的内容(如文本、流媒体以及可视化对象)中获取结构化的具有语义的知识。知识获取的关键问题有:如何将隐性的知识显性化;如何从不同的来源中获取并集成知识(如来源于 WWW 的分布式的资源);如何从不同的媒体中(如从图表和非结构化文本中)获取知识等。

而知识建模是知识技术中的关键技术,它是关联知识获取和知识利用的桥梁。知识模型的结构一方面需要能够容纳所获取的知识,另一方面需要呈现相关的知识,以使能够被应用于问题解决环境。知识建模根据需求的不同有着不同的研究方向,一个重要的知识建模领域就是本体,通过它可以定义一个领域内的通用概念、属性、关系以及相应的公理。

知识技术的知识获取和知识建模能够形成形式化的知识,知识技术的研究者希望在此基础之上,改变手工建设知识库的现实状况,实现知识库的自动建设。在知识库的基础之上,可以采用或拓展 KDD 的相关技术,实现更深意义上的知识发现。

7.6.2 e-Science环境下的文本挖掘技术

在 e-Science 界,越来越多的人都意识到,需要更好的方法来处理不断产生的大量的联机研究文献,而自动化文本处理技术的所取的成就,也使人们看到了解决问题的希望。文本挖掘技术的目标是利用语言技术和知识发现技术,从结构化或非结构化的文本中抽取新的、有用的知识。e-Science 网格环境的出现,为大规模文本挖掘系统的构建和利用提供了可能。

1. CLEF及其建设目标

CLEF (A Co-operative Clinical e-Science Framework) 的目标是,在充分利用网格技术的基础之上,在充分考虑患者为保护个人隐私的需求同时,将门诊信息进行信息获取、信息整合、信息标注,并嵌入到实际的临床系统中应用。它将构建如下环境:

(1) 一个共享的知识仓储,可以将门诊、图像和基因信息集成在一起,能够与医学记录、文献和 Web 资源进行连接。

(2) 信息获取及语言工具,可以对门诊信息进行知识抽取,添加描述信息,并且将这信息植入到仓储中。

(3) 信息集成工具。将新产生的医学记录和仓储中的医学记录,以及其他孤立的各种研究信息有机集成在一起。

(4) 信息展示工具,使这些信息更加方便地获取。

(5) 一个知识管理环境。利用这一工作环境,可以增进生物医学的研究。

整个 CLEF 的工作流程如图 7-37 所示。

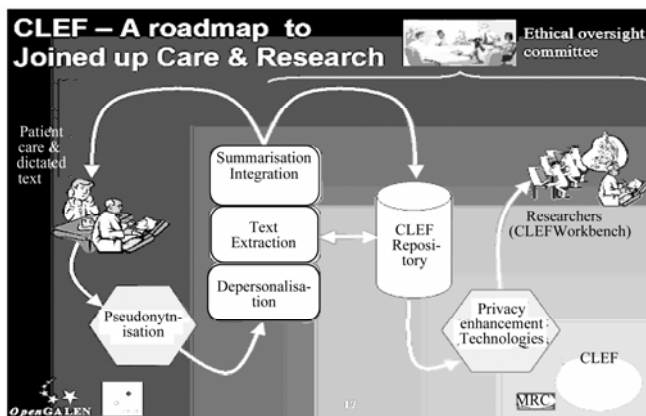


图 7-37 CLEF的工作流程¹

针对门诊报告（如病历概述、放射报告和病理学报告）大多都以文本形式存在的情况，需要利用到信息抽取技术实现对医学记录中关键信息的抽取。

为了实现 CLEF 中信息的抽取，谢菲尔德大学在 MyGrid 的基础之上，开发了一个名为 AMBIT（Acquiring Medical and Biological Information from Text，从文本之中获取医学和生物信息）的系统。这一系统的构成如图 7-38 所示。

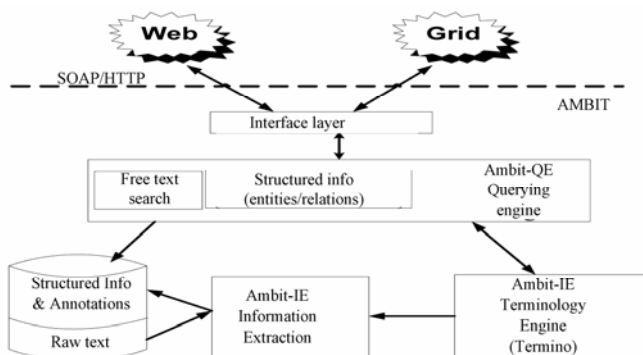


图 7-38 AMBIT²

2. CLEF的系统组件

该系统由以下几个组件组成：

（1）接口层：系统提供基于 SOAP 的 Web 服务接口，可以实现与其他系统的数据交换和集成。目前系统希望提供支持 OGSA-DAI 的网格接口。

¹ CLEF: Clinical e-Science Framework,[EB/OL]. (2008-8-20), <http://www.clinical-escience.org/>

² AMBIT: Acquiring Medical and Biological Information from Text[EB/OL]. (2008-8-20), [http:// www.nesc.ac.uk/events/ahm2003/AHMCD/pdf/090.pdf](http://www.nesc.ac.uk/events/ahm2003/AHMCD/pdf/090.pdf)

(2) 查询引擎 (Query Engine): 查询引擎允许用户查询数据库、自由文本信息和其他相关信息。

(3) 原始和被标注文本数据库: 原始的和被标注的文本信息都被存储在关系数据库的表中。

(4) 信息抽取引擎 (Information Extraction Engine): 信息抽取引擎在原始文本中抽取关键实体及其相关关系, 并将其作为一系列的结构化对象存储在数据库。

(5) 术语引擎 (Terminology Engine): 术语引擎一方面为信息抽取引擎标识出文本中的技术名词; 另一方面为查询引擎提供相应的术语查询支持。

7.6.3 基于网格的知识发现服务技术

Discovery Net 项目是英国一个对各种传感器上收集的科学数据进行知识发现的服务框架。Discovery Net 项目的研究者认为, e-Science 平台与网格计算机平台相比, 是一种更高层次的应用平台, 其目标是在所有设备、传感器、数据库、分析组件和计算资源都可以被网格连接的情况下, 能够让终端科学家获取新的知识。

Discovery Net 系统通过工作流和管道方式来处理科学数据的分析。在这一个框架之中, 各个服务被设计成为一个输入和输出接口可见的黑盒子。这些服务可以被连接起来成为一个序列 (或操作过程)。Discovery Net 框架提供了一个开放的数据挖掘平台, 能够在知识发现的过程中集成分布式的数据资源和分布式的挖掘工具。Discovery Net 的结构如图 7-39 所示。

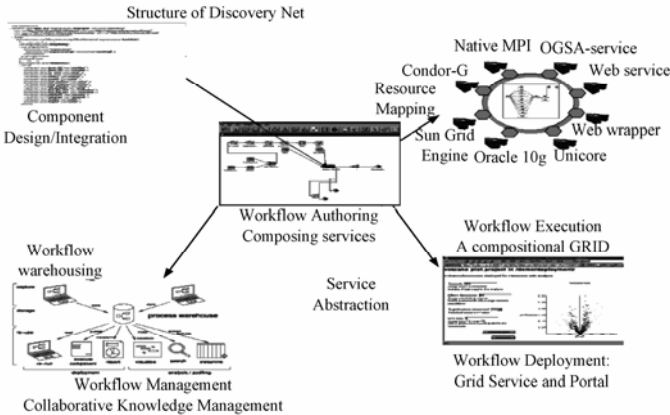


图 7-39 Discovery Net¹

Discovery Net 主要工作流程如下:

(1) 用户通过 Discovery Net 的工作流定义工具可以定义他们的数据分析工作流。

¹ Vasa Curcin, SARS Analysis on the Grid: Discovery Net in Bioinformatics [EB/OL]. (2008-8-20), <http://www.nesc.ac.uk/events/ahm2004/presentations/80.ppt>

在此阶段，用户可以动态地浏览、集成各种数据服务和分析服务，并且将它们动态地组合成一个工作流。各种数据服务和分析服务基于标准的接口定义，除了支持它本身的接口之外，还支持 Web 服务接口。

(2) 编辑好的工作流被发送给工作流执行引擎，工作流引擎利用了多种方法实现各种资源的定位和数据传输，通过在不同服务器上的分布式运行。

(3) 通过利用 Discovery Net InfoGrid 基础设施，用户可以动态地在他们的工作流中存取和集成各种类型的异构数据集。

(4) 通过 Discovery Net Deployment Tool，用户可以对它们工作流进行抽象并且封装成为一个新的服务进行发布，使更多的用户可以通过 Web 服务或其他接口使用这一服务。

(5) Discovery Net 提供了一个工作流数据仓库，可以存储用户定义的工作流，用户可以对 Discovery Net 中的工作流进行查询。

在 Discovery Net 中，主要的技术包括，网络和分布式计算技术、工作流和服务创建技术、数据挖掘和可视化技术、数据存取和信息结构化技术、实时的高吞吐量监测设备。它们在 Discovery Net 中的结构关系如图 7-40 所示。

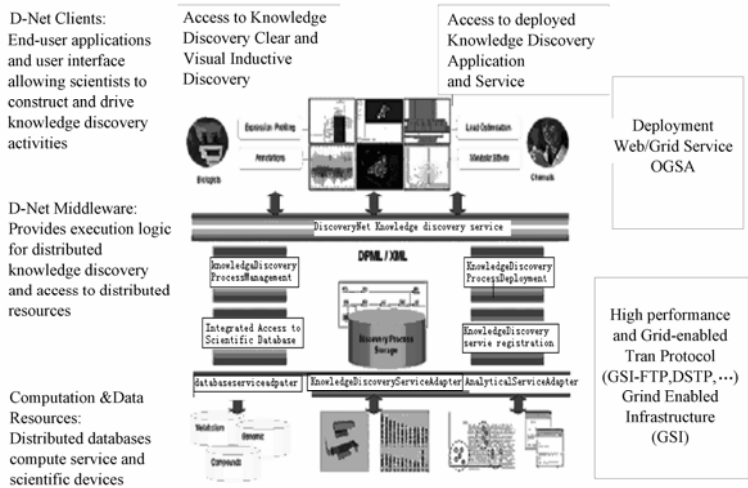


图 7-40 Discovery Net 的技术结构关系¹

Discovery Net 主要被应用于生命科学、实时环境监测和地质灾害建模。在 2002 年 IEEE SC2002 超级计算机会议上，Discovery Net 与 Delta Dot 公司提供的实时 DNA 测序平台一起工作，完成了实时的基因测序和标识工作。Discovery Net 还进行了 SARS 病毒的变异过程的重构和构建 SARS 的生物学、市区污染监测和地震等研究。

¹ Patrick Wendel, Discovery Net: A UK e-Science Pilot Project for Grid-based Knowledge Discovery Services, [EB/OL]. (2008-8-20), <http://static.msi.umn.edu/general/Symposia/dmem/wendel.pdf>

第 8 章 研究对象的建模和仿真技术

8.1 可视化技术

可视化通过图形化的方式来呈现数据。可视化已经成为 e-Science 的核心功能，成为许多 e-Science 项目和网格项目的重要组成部分。在科学研究中，对于采集、接收、观测、仿真获得的大量数据，采用了可视化技术，能够为科研人员提供更加直观、形象的手段了解相关的事实和数据真相。

8.1.1 基于网格的可视化技术框架

传统的可视化方法是在用户的电脑上使用一个独立的软件来运行可视化，通过网格，可以在一个分布式的、异构的环境将硬件和软件进行集成。计算操控（Computational Steering）将模拟仿真和可视化进行紧密结合，已经成为当前有效使用网格资源的重要范式。可视化参考模型如图 8-1 所示。

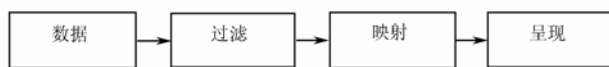


图 8-1 Haber and McNabb参考模型¹

大量的基于数据流范式的互动可视化系统均基于该参考模型，包括 Gviz 项目中的 IRIS EXPLORER 系统。在该模型中，原始的可视化数据通过程序，进行过滤，转换形成一个抽象的几何表示，最终成为一个几何图形，实现可视化。

基于该模型的可视化系统的最重要的两个特点是：开放性和可扩展性。

开放性表现在两个方面。一方面，大量的标准化的模块分布在系统中；另一方面，用户可以通过封装他们自己的编码作为模块，增加到这个库中，同时可以像任何标准模块一样装载到系统中；

可扩展性也表现在两个方面，一方面允许系统在此框架之下，随着计算机技术的不断进步，而不断扩展新的模块，改进原有的算法；另一方面，可扩展性还表现在 e-Science 环境中，能够生成特定的协作模块，用于链接处于不同地点的用户执行的程序。这一功能在 EPSRC（The Engineering and Physical Sciences Research Council，英国工程和自然科学研究委员会）COVISA 项目中的 IRIS EXPLORER 就已经实现，现在已经成为这个系统的一部分。

¹ Visualisation on the Grid: A Web Service Approach, Stuart M. Charters, [EB/OL]. (2008-8-20),<http://www.allhands.org.uk/2004/proceedings/proceedings/proceedings.pdf>

模块化可视化环境（Modular Visualization Environments，MVEs），如 IRIS EXPLORER 和 AVS/Express，可以让用户采用最简单易用的作业方式，而不用编写程序代码就可以有效地开发 3D 可视化应用系统；AVS/Express 是 AVS 公司开发的世界领先的多维可视化开发平台，可在各种操作系统下开发可视化应用程序，利用它可以快速建立具有交互式可视化和图形功能的科学和商业应用程序。开发者可以使用其面向对象的可视化编程环境，在一个开放和可扩展的环境下快速建立应用程序原型，处理海量数据。AVS/Express 提供了有关图形、图像、数据可视化、数据库接口、注释和硬拷贝输出等方面的许多先进技术。

在模块可视化环境下，可视化的运行要求运行程序的机器必须具有充足的内存、处理器和强大的图形处理能力。

1. 基于网络可视化技术框架

在传统可视化参考模型的基础上，Stuart M. Charters等提出了基于网络的可视化计算操控框架¹，如图 8-2 所示。

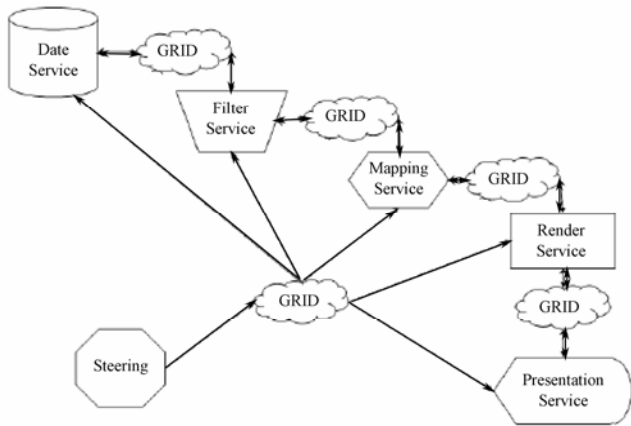


图 8-2 基于网络的可视化技术框架

基于网络的可视化计算操控框架有以下特点：

- （1）主要由数据服务、过滤服务、映射服务、表示服务和呈现服务五个服务组成一个工作链；
- （2）工作链中的每一个服务都是一种网格服务；
- （3）每一个服务都可以在不同的机器上运行；
- （4）每一个服务都可以独立进行升级。

在这个工作链中，每一个部分都维持传统的数据前向流。用户可以对框架进行

¹ Visualisation on the Grid: A Web Service Approach, Stuart M. Charters, [EB/OL]. (2008-8-20), <http://www.allhands.org.uk/2004/proceedings/proceedings/proceedings.pdf>

修改、控制和执行整个工作链。与其他基于 C/S 模式的网格可视化不同，Stuart M. Charters 提出的这个框架可能将分布在不同地理位置的相关可视化网格服务集成到这一框架之内。这种可视化模式允许可视化的每个部分在不同的机器或者具有不同计算能力、存储能力和图形处理能力的机器上运行，通过这种方式，可以最大利用网格资源，降低总成本，增加可视化的灵活性。

模块可视化环境包括可视化设计工具、可视化执行环境和终端用户的显示。这些都包含在一个单个的软件环境中。

2. 基于网格的可视化技术框架的主要优点

基于网格的可视化技术框架有四方面的优点。

(1) 工作链中每一服务可以部署到适当的机器上运行

工作链上的每一个服务都允许在不同的地理位置不同的机器上运行，这种独立性允许每种服务都可以根据其被执行的方式，以最合适的方式进行部署。例如，呈现服务可以在一个高效的图形工作站上运行以生成高质量的图片，映射服务可以通过多个并行计算机来实现数据转换，而过滤服务部署在具有高效存储能力的计算机上运行。

(2) 可以实现可视化设计人员与可视化用户之间的独立工作

在该框架的基础上，可以实现一个用户在某一时间、某一地点建立或者构建可视化服务，而另外一个用户可以在不同的时间、不同的地点充分利用这些可视化服务。这个特点非常有利于异地科学家之间的协作。

(3) 增强协作

设计上的灵活性，允许在工作链上的每一个服务进行灵活分布，使得不同团体之间，可以同时在不同地点或者不同时间不同地点之间方便地实现协作。不同地点的用户可以运行不同的表示服务以支持他们特殊的显示装置，相反地，一个表示服务可以支持不同的呈现服务。

(4) 可以在运行时刻根据具体情况，生成合适的图像

这一框架使用元数据来描述各层服务，元数据可以从相反的方向传向工作链，例如可以将呈现层服务的元数据传向表示层服务，当表示层服务接收到呈现层传递给它的反映呈现条件的元数据后，表示层服务能够选择最适合呈现层环境的表示服务来支持呈现服务，从而根据运行时刻的具体情况，生成合适的图像。

3. 基于网格的可视化技术框架具体特点

基于网格的可视化技术框架的特点包括以下三点。

(1) 元数据与网格服务数据

基于网格的可视化计算操控框架为每个服务都提供了网格服务数据项，在这个网格服务数据项中，可以提供描述这一服务的元数据，让其他服务和用户可以使用这些元数据以了解如何使用这个可视化服务。这一机制允许不同层的可视化服务能

够自身情况调用和配置不同的服务。同时，网格可视化框架还提供了一些检查机制，避免不同的数据类型或不同的服务被错误地组合。服务数据还可提供更多关于服务能力方面的细节信息，通过这些信息能够帮助不同的服务在服务数量、安全、信任和隐私等方面进行更好的沟通。

（2）组件化的可视化框架

该框架允许组合多种可视化组件，也支持可视化组件的拆解（decoupling），这样可以使一个可视化专家为在世界上任何地方的最终用户构建可视化环境。利用网格服务建立可视化，可视化组件将持续保留在网格上，可以让可视化专家在不同的时间、不同的地点下实现可视化的协作构建。

可视化的远程组合和配置，可以让可视化专家更加有效地进行可视化开发，并且更加高效地利用他们的时间，使这些专家能够将精力集中在他所应当做的事上。

（3）操控/最终用户控制

可视化的最终用户要求能够对可视化工作链中的各个组件实行操控，为此，可视化工作链上的每一阶段都要求提供操控功能。这意味着这些接口可以从呈现服务中分离出来，在可视化过程中可以不存在，而只需在要用到的时候及时调用即可。

可视化的控制机制与用于进行可视化操控的所有的工具相协调，也与操控界面的工作链相协调。

8.1.2 GViz分析

1. 项目简介

GViz 项目（Visualisation and Steering for the Grid）成功地将 IRIS EXPLORER “网格化”。在这个项目中，可视化被表达为一个工作链，在数据读取的基础之上，通过几何方式构建可视化，最终将几何表现为可视化。

2. 项目目标

GViz是英国e-Science核心程序项目。该项目的目标是研究e-Science的可视化中间件，即研究在网格框架中，现有的可视化系统改变为可操控制系统的程度，希望最终能够通过扩展IRIS EXPLORER生成新的模块¹。通过这个项目，初步实现如何进一步将已有的可视化系统升级为网格系统，如何将可视化环境和仿真模拟环境进行链接。

3. 参与机构

参与本项目的机构主要有学术机构和一些商业公司。其中，学术机构包括利兹大学，牛津大学，牛津布鲁克斯大学和英国科研委员会中心实验室；主要公司包括：

¹ Jason Wood, Ken Brodrie, gViz – Visualization and Steering for the Grid [EB/OL]. (2008-8-20), <http://www.nesc.ac.uk/events/ahm2003/AHMCD/pdf/030.pdf>

Streamline computing 公司，NAG 公司和英国 IBM 公司。

4. 技术原理

在该项目中定义了一个新的 XML 应用，称为 skML，能够独立地描述可视化资源。该语言支持协作可视化，允许以文档的方式，将相关概念在不同的软件系统之间进行传递。作为 skML 语言的补充，GViz 项目开发了一个编辑器(SVG Map Editor)，以实现数据流网络的编辑。SVG Map Editor 的用户界面以 IRIS Explorer 的编辑器为模版，与 AVS 也很类似。SVG Map Editor 使用了 Web 技术、SVG (Scalable Vector Graphics) 和 JavaScript，因此可以在网络浏览器上运行。

GViz 项目开发了 gViz 计算操控类库(GViz Computational Steering Library)，gViz 计算操控类库提供接口可实现各种可视化应用。gViz 计算操控类库又分为两个部分，一部分为科学家所用，通过它的 API，科学家可以装载他们的模拟数据；另一部分类库，提供自动匹配能力，可以在最终的可视化系统中实现相关部件的集成。GViz 库设计的框架如图 8-3 所示。

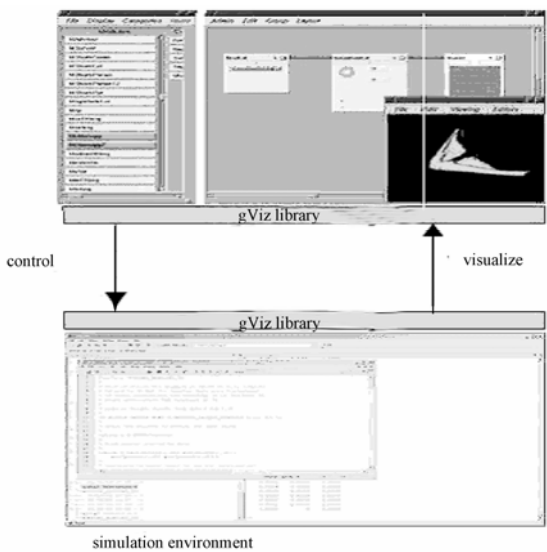


图 8-3 可视化环境¹

操控参数从前端系统（在此指 IRIS EXPLORER）中输入，并通过 gViz 软件，传输到底层的模拟仿真环境中，底层模拟仿真得到的结果又通过 gViz 软件，返还给前端系统。

¹ Ken Brodlie, Jason Wood, David Duce, Musbah Sagar, gViz: Visualization and Computational Steering on the Grid [EB/OL]. (2008-8-20),<http://www.nesc.ac.uk/events/ahm2004/presentations/67.ppt>

8.1.3 e-Demand分析

e-Demand 项目是关于构建立体可视化的网格服务框架，主要目标是基于网格服务在分布式环境中来支持立体可视化。其框架允许可视化工作链上的每一个部分以网格服务的方式实现，并通过点对点的方式来利用，以获得高度的灵活性和可配置性，确保立体可视化能够利用与任务相关的最佳资源。

可视化框架的当前实施情况已经从传统的可视化链转向了一个分布式的服务框架。图 8-10 显示了 OGSA 环境中的服务的一个例子。开发人员负责开发作为框架中某个可视化服务的接口，撰写 WSDL（Web Service Description Language，Web 服务描述语言）文档，实现具体的服务和生成器。

e-Demand 项目利用了 Java 语言开发，并利用了 OGSA 工具包。利用了 WSDL 发布每个服务的接口。每个 WSDL 接口都有具体的 Java 接口实现。Java 接口至少提供两个方法：一个方法用于设置数据源，通过这个数据源，服务可以获取数据；另一个方法允许其他的服务可以从这个服务上获取数据。Service 框架是一个抽象的网格服务，它包括很多的模版样例，可以非常方便地构建任务所需的具体服务实现。对于每一个服务，都有一个服务 factory，它负责在某个主机上，为某个服务“制造”一个实例。

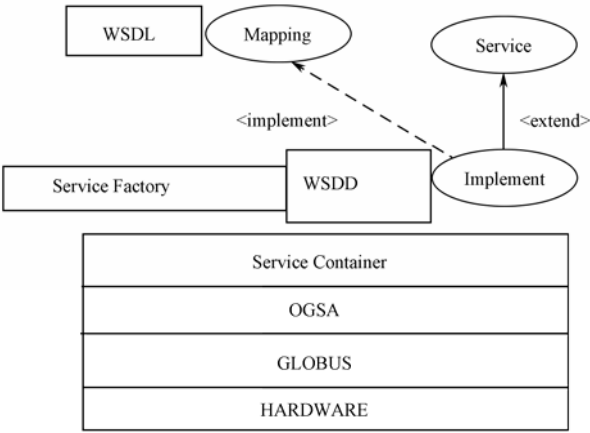


图 8-4 OGSA环境中的e-Demand服务¹

WSDD（Web Services Deployment Descriptor，Web 服务部署描述符）将服务制造工厂、WSDL 和和特定的服务在服务容器中关联在一起。对于服务而言，服务容

¹ Stuart M. Charters, Nicolas S. Holliman & Malcolm Munro, Visualisation in e-Demand: A Grid Service Architecture for Visualisation,[EB/OL]. (2008-8-20),<http://www.nesc.ac.uk/events/ahm2003/presentations/14.ppt>

器就是一个服务的运行环境，它负责对其他的软件和物理资源进行管理，并提供相应的安全措施。

服务间的通信利用了 SOAP (Simple Object Access Protocol, 简单对象访问协议) 协议，它在运行在 HTTP 等协议的基础之上。

e-Demand 项目的研究成果已经被应用到了化学领域，具体地讲是构建了一个 X 射线衍射结晶学的实例 (如图 8-5)，以演示和测试这一技术框架。这一应用可以将一些晶体分析的数据，转换成为一个由球和棒代表的分子，以表示分子中存在的电磁强度。

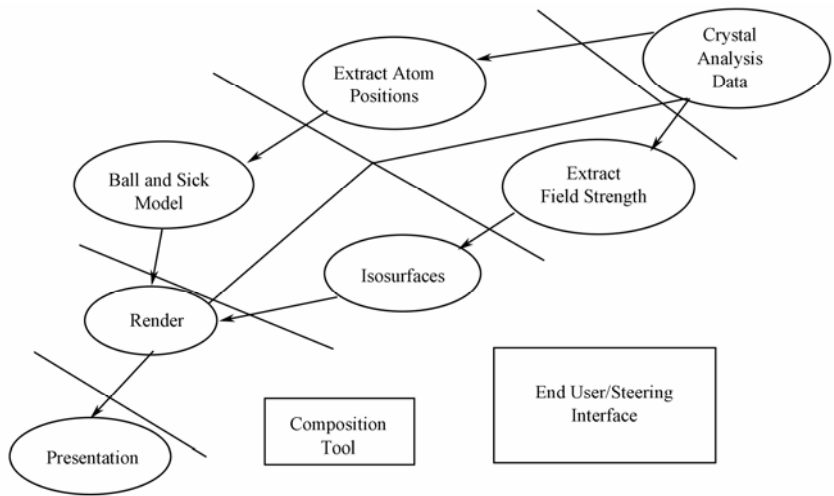


图 8-5 e-Demand化学实例¹

8.1.4 飞行器中的电磁散射

电磁散射在飞行器设计和遥测中起着十分重要的作用。可视化工具可以交互式地检测结构表面上的散射模式，检测雷达截面 (Radar Cross Section, RCS)，以及在三维空间中检测复杂结构中的电磁场情况。

由于门户具有用户友好、独立于平台和不受地域限制、可以方便呈现磁场散射数据的特点，剑桥大学应用数学和理论物理系的波研究组提出了基于门户方式，通过协同工作，开发复杂结构中的电磁场散射可视化技术的思路。

飞行器中的电磁散射项目组以基于 Grid Portal Toolkit (Gridport) 的剑桥大学的

¹ Stuart M. Charters, Nicolas S. Holliman and Malcolm Munro, Visualisation in e-Demand: A Grid Service Architecture for Stereoscopic Visualisation, [EB/OL]. (2008-8-20), <http://www.nesc.ac.uk/events/ahm2003/AHMCD/pdf/161.pdf>

EMS 网络门户为基础进行开发。利用了开源工具 Visualization Toolkit 以可视化展示每一次电磁散射计算之后的复杂数据。

研究组利用的 VTK (Visualization Toolkit, 可视化工具集) 是一个 C++ 类库, 它直接集成了许多可视化算法。VTK 的框架使得复杂的三维图形应用变得相对简单, 并且可以运用面向对象的技术。许多可视化算法的实现为更深入的可视化应用提供了一个良好的基础平台。

在技术上, 研究组对每个空间上的采样点都有一个复杂的三维向量, 提供振幅或者强度形式的标量值, 这些标量数据被映射到颜色查找表中, 大量过滤器和映射器的使用创建了一个可以渲染的几何图形。

表面电流和内部磁场的数据在本质上都是同一种数据结构的数据, 这样就简化了用于可视化的数据装载类。实际上, 同一个装载类可以同时用于表面电流数据和内部磁场数据。一旦数据被装载, 它被存储在基于 VTK 的队列中, 在 VTK 队列中, 提供了些更高层次的功能, 比如边界检测和动态的队列大小确定。

对象的可视化外观取决于对相关数据的拓扑和几何表示方法。比如, 数据可以绘制成单个的带有标量属性的数据点。其他的类型的数据, 比如说曲线和四边形, 可以通过插入标量获得更好的效果。三维数据类型可以最好的代表电磁场, 起初使用的是 voxel (在地理信息系统的真三维空间中, 用以进行空间信息的数据记录、处理、表示等所采用的具有一定大小的最小体积单元) 数据类型, 它能给出令人满意的视觉效果, 但是它还需要进行更多的验证, 特别是利用高性能硬件上的 VTK 体渲染类。

研究组认为, 他们将来的工作将包括增加新的图形用户接口, 他们也正在评估是否使用 wxWindows 工具箱。同 VTK 一样, wxWindows 是拥有众多用户支持的开源项目, 而且 wxWindows 是跨平台的产品, 它能轻易地移植到其他操作系统, 以满足日益增长的需求。

8.2 虚拟观测台技术

虚拟观测台 (the Virtual Observation, VO) 技术的目的是实现对现有的空间、观测台、天体测量数据等天文学数据档案的全球通用访问。虚拟观测台提供通用而标准访问方式、高速的宽带接入、以及很多的分析工具, 以构建协同的科研合作实体。虚拟观测台是一种框架, 通过它可以推动数据存档和数据挖掘协议的标准化, 从而进一步实现对各种不同数据的全球访问。现在的虚拟观测技术主要集中在欧洲、南北美、亚洲、大洋洲等地的天文团体的全球合作上。

8.2.1 VO概念

在虚拟观测台出现之前，在天文学家们处理信息的时候，为了将不同的观测结果组合起来，他们不得不将不同图片中的坐标系统一，然后再进行匹配。这就是一项很困难也很耗时间的工作。在科学信息可以被提取出来之前，天文学家还需要对观测结果进行清洁、校准和分析。不同的望远镜和仪器都会生成自己特有的数据，都需要清理出去。在进行检索的时候，根据天文学家们的兴趣不同，他们可能会为这些对象编制一个目录，也可能会将自己所发现的东西和现有期刊或者因特网上的目录作一个对比。这同样又是一项艰苦的工作。这里有成千上万个目录，每个又对应着百万级的对象。

VO 可以实现以下的目标：在众多的观测记录中找到所需要的对象，或者对象集，并且能够在其中自由切换；能对多种观测结果进行匹配；对某一个对象的全世界的目录进行比较；对不同时间得到的观测记录进行比较；突破仪器的限制。同时还可以从 VO 中得到如下的科学结论：更好理解潜在的物理显示；可以从标准化后的特性进行合并的过程中产生的信息中发现新的对象；可以发现在天空中不断移动的新的物体，比如说小行星；发现更多的恒星甚至是超新星；最重要的是，利用新的方法来合并数据的时候，也许会有意想不到的发现。

虚拟观测台会像因特网一样便捷，为天文学家提供专门的数据。通过对大型计算机和数据中心的访问，可以将满含了结构化和标准化后的信息的实例库推送到天文学家的桌面上来，以供科学研究。

从本质上说，VO 就是一套支持互操作和协同操作的软件系统，它支持用户通过一种无缝的透明的方式来对多个信息中心进行查询。同时 VO 提供强大的最新的分析工具和可视化工具，为数据中心和数据提供者一个标准的框架，利用它们的信息进行发布和传递等服务。

8.2.2 AstroGrid项目

AstroGrid 是英国对 VO（Virtual Observation）领域的一个巨大贡献，推动了 VO 的快速发展。它是一个由英国主要天文数据中心组成的联盟。这个项目主要是由 PPARC 提供基金，他们提供了 370 万英镑，但是也通过 AVO（Astrophysical Virtual Observatory）合作得到了欧盟的一些支持。AstroGrid 的资金去向相当广泛，包括太阳物理和太阳陆地（空间等离子体）物理，还有光学、无线电和 X 射线天文学。第一阶段的研究工作从 2001 年 9 月开始到 2002 年 12 月结束，此外还与国际伙伴对新的标准进行了首次的开发。AstroGrid 的第二阶段——建设阶段已于 2003 年 1 月启动。作为 OGSA-DAI 软件的早期采纳者，AstroGrid 已经成为英国 e-Science 项目的有机组成部分。

全世界范围内的其他 VO 项目也开始兴起，国际虚拟观测台联盟（International Virtual Observatory Alliance, IVOA）的重要性也在日渐增加。最初它是一个高级管理讨论论坛，不久就开展了一组技术问题的工作组，以及一系列的专题会议，旨在对数据、元数据以及软件模块互操作性制订通用国际化标准。

AstroGrid 的目标是，提高联机天文学研究的质量、效率、简单性和速度；使得来自不同数据源的比较和整合变得无缝及透明；把数据分析壁垒留给边缘学科研究；使得涉及一对大量数据集进行操作的科学研究尽可能的简单和有功效。

为了达到这些目标，AstroGrid 制订了一系列的措施：与 IVOA 伙伴一起为数据、元数据、数据交换和出处开发一套标准；为数据服务开发一个软件基础结构；建立一个可以被 AstroGrid 和主要数据中心共享的资源物理网络；建立并维护一个 AstroGrid 服务和资源注册机制；实现一个基于主要的 UK 数据库的工作虚拟观测台系统，并为这个系统提供一个用户界面；通过构建或应用为 VO 系统提供一组科学用户工具；在 VO 工作中为英国建立一个领先的位置。AstroGrid 的架构如图 8-6 所示。

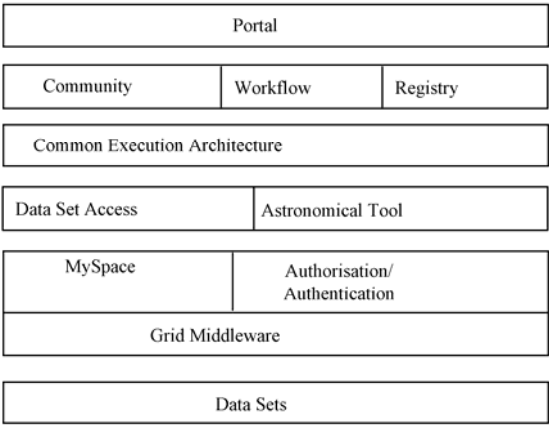


图 8-6 AstroGrid架构¹

① 门户（Portal）：这是一个基于服务器的组件，提供了一个对服务进行访问的界面。在 AstroGrid 中所有的需要与用户进行交互的组件都会通过这个门户。这个门户提供了一个模块化的很灵活的方式，开发者可以根据自己的需要来订对它进行修改。

② 社区（Community）：通过它可以在一个工作组中建立一个灵活的社区。当一个社区的管理员获得授权的时候，他就可以为这个社区中的个人进行权限的分配。

③ 工作流（WorkFlow）：通过这个可以建立复杂的任务，比如说查询、数据分析，上传下载数据或者利用表格或图片的方式输出数据。

¹ AstroGrid as an e-Learning Environment [EB/OL]. (2008-8-20), http://hal.archives-ouvertes.fr/docs/00/19/05/88/PDF/Contreras-P_AstroGrid_2005.pdf

- ④ 通用可执行架构 (Common Execution Architecture): 是一组应用组件抽象。
- ⑤ 数据集存取 (DataSet Access): 是一组界面组件, 可以获取标准的查询, 将它翻译成数据集并且执行, 然后将结果返回给数据中心。
- ⑥ 天文学工具 (Astronomical Tool): 是一套重要的天文学工具, 包括对象目录创建器 SExtractor 和广度红移分析器 HyperZ。这些都通过门户同 AstroGrid 集成在一起。
- ⑦ MySpace: 它定义了一个虚拟的空间, 将临时的或者长期的所有数据都定位在这里。
- ⑧ 认证与授权 (Authorisation/Authentication): 是一套管理身份认证和存储权限的组件, 这也组成了 AstroGrid 中的安全机制。
- ⑨ 网格中间件 (Grid MiddleWare): 是 AstroGrid 同其他天文数据中心整合以及共享数据的通道。
- ⑩ 数据集 (Data Sets): 是分布在不同的数据中心的数据的集合。

AstroGrid 在客户端提供了两个很重要的可视化工具, 其一是 MovieMaker, 它的作用就是让用户利用现有资源制作一个在一段时间内太阳活动情况的视频。它既可以通过客户端工作区使用, 也可以通过门户使用。另一个就是 RedShift Maker (图 8-7), 它可以从光学图片数据中演绎出光度红移。

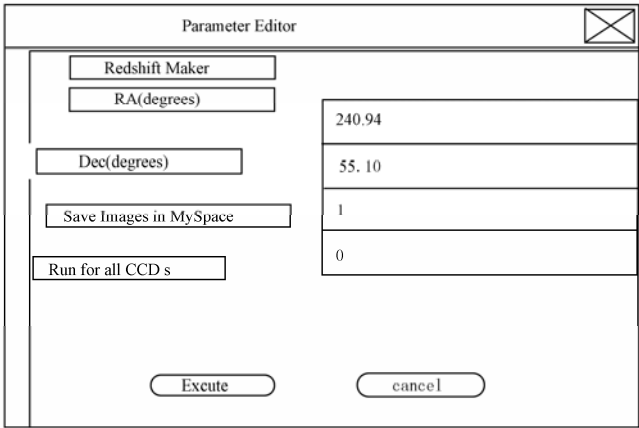


图 8-7 RedShift Maker 参数输入界面

8.3 计算机动画技术

网格计算机动画技术是建立在网格可视化技术以及网格数据传输技术之上的一种应用, 目的在于解决大规模分布式环境下的海量数据的计算机动画相关问题。目前 e-Science 项目中涉及计算机动画的项目有 The PGPGrid Project。该项目的目的在于探索计算机动画在虚拟组织中的发展前景, 通过网格中间件对三个异地的不同组织之间的数据和资源进行协调, 最终生成一个短篇计算机动画。

8.3.1 计算机动画技术研究

1. 计算机动画的定义

计算机动画是指采用图形与图像的处理技术,借助于编程或动画制作软件生成一系列的景物画面,其中当前帧是前一帧的部分修改。所谓动画也就使一幅图像“活”起来的过程。使用动画可以清楚的表现出一个事件的过程,或是展现一个活灵活现的画面。动画是一门通过在连续多格的胶片上拍摄一系列单个画面,从而产生动态视觉的技术和艺术,这种视觉是通过将胶片以一定的数率放映体现出来的。实验证明:动画和电影的画面刷新率为 24 帧/秒,即每秒放映 24 幅画面,则人眼看到的是连续的画面效果。计算机动画是采用连续播放静止图像的方法产生物体运动的效果。

2. 基于网格的计算机动画

基于网格的计算机动画实际主要利用网格作为中间件,充分利用网格上的资源与数据,实现异地的数据快速准确传输,从而协作完成计算机动画。在构成动画本身的技术上和传统计算机动画没有太大差异,它们的差异主要在于制作计算机动画的整个过程中分布式异地数据和资源的传输和协调。基于网格的计算机动画技术能够更好地利用网格上的计算资源,更快更准确地获得异地的采集数据,从而生成更高质量的计算机动画。

3. 计算机动画的工作原理:关键帧动画

关键帧的概念来源于传统的卡通片制作。在早期 Walt Disney 的制作室中,熟练的动画师设计卡通片中的关键画面,也即所谓的关键帧,然后由一般的动画师设计中间帧。在三维计算机动画中,中间帧的生成由计算机来完成,插值代替了设计中间帧的动画师。所有影响画面图象的参数都可成为关键帧的参数,如位置、旋转角、纹理的参数等。关键帧技术是计算机动画中最基本并且运用最广泛的方法。另外一种动画设置方法是样条驱动动画。在这种方法中,用户采用交互方式指定物体运动的轨迹样条。几乎所有的动画软件如 Alias、Softimage、Wavefront、TDI、3DS 等都提供这两种基本的动画设置方法。

无论是样条驱动动画还是关键帧插值方法,都会碰到这个问题:给定物体运动的轨迹,求物体在某一帧画面中的位置。物体运动的轨迹一般由参数样条来表示。如果直接将参数和帧频联系起来,对参数空间进行等间隔采样,有可能带来运动的不均匀性。为了使物体沿一样条匀速运动,必须建立弧长与样条参数的一一对应关系。Guenter 等提出用 Gauss 型数值积分方法计算弧长,用 Newton-Raphason 迭代来确定给定弧长点在曲线上的位置,并采用查找表法记录参数点弧长值的方法来加速计算。在动画设计中,动画师经常需调整物体运动的轨迹来观察物体运动的效果,交互的速度是一个很重要的因素。Watt 等提出了用向前差分加查找表的方法来提高交互的速度。在精度要求不是很高的情况下,他们的方法非常有效。

从原理上讲, 关键帧插值问题可归结为参数插值问题, 传统的插值方法都可应用到关键帧方法中。但关键帧插值又与纯数学的插值不同, 它有其特殊性。一个好的关键帧插值方法必须能够产生逼真的运动效果并能给用户提供方便有效的控制手段。一个特定的运动从空间轨迹来看可能是正确的, 但从运动学或动画设计来看可能是错误的或者不合适的。用户必须能够控制运动的运动学特性, 即通过调整插值函数来改变运动的速度和加速度。为了很好地解决插值过程中的时间控制问题, Steketee 等提出了用双插值的方法来控制运动参数。其中之一为位置样条, 它是位置对关键帧的函数; 另一条为运动样条, 它是关键帧对时间的函数。Kochanek 等提出了一类适合于 keyframe 系统的三次插值样条, 他们把关键帧处的切矢量分成入矢量和出矢量两部分, 并引入三个参数: 张量 t 、连续量 c 和偏移量 b 对样条进行控制。该方法已在许多动画系统中得到了应用。

关键帧插值系统中要解决的另一个问题是物体朝向的插值问题。物体的朝向一般可由 Euler 角来表示, 因此朝向的插值问题可简单地转化为三个欧拉角的插值问题, 但欧拉角表示又有它的局限性。因为旋转矩阵是不可交换的, 基于欧拉角的旋转一定要按某个特定的次序进行; 此外, 等量的欧拉角变化不一定引起等量的旋转变化, 从而导致旋转的不均匀; 欧拉角还有可能导致自由度的丧失, 即所谓的“框锁定”(gimble lock) 现象。Shoemake 为了解决因采用欧拉角表示引起的缺陷, 最早把四元数引入到动画中, 并提出了用单位四元数空间上的 Bezier 样条来插值四元数。Barr 等人提出了一个采用四元数对带有角速度约束的景物的朝向进行光滑插值方法, 他们的方法允许用户对轨迹端点处的角速度进行约束。Kim 通过构造一组新的基, 提出了把空间曲线变化到单位四元数空间曲线的一般性方法。

4. 计算机动画发展里程碑

20 世纪 60 年代, 二维计算机辅助动画系统诞生, 比如加拿大的 MSGEN 系统, 美国的 CAAS 系统 (Computer-assisted animation system, 计算机辅助动画系统), 大大提高了二维动画的质量和制作效率。70 年代, 三维图形与动画的基本技术被开发, 其间一小批领导三维动画与图像的公司出现, 一些三维可明暗着色系统被完成。80 年代的主要发展在于优化了 70 年代出现的模型和阴影技术, 比如康奈尔大学的辐射度方法, JPL (Jet Propulsion Laboratory) 实验室的运动动态, 加利福尼亚大学的样条模型, 多伦多大学的过程技术, 俄亥俄州立大学的人物分级动画与逆向运动学, 蒙特利尔大学的人物动画与嘴唇同步, 东京大学的气泡表面模型技术, 广岛大学的辐射度与灯光。90 年代, 动力学仿真技术、三维仿真演员系统和自主动画出现。21 世纪初, 在 e-Science 的发展下, 出来了基于网格技术的计算机动画, 为分布式的计算机动画提供了技术支持。

5. 计算机动画的二维与三维

计算机动画从视觉效果上可以分为二维动画和三维动画。二维动画即平面上的

画面，在纸张、照片或计算机屏幕显示，无论画面的立体感多强，终究是二维空间上模拟真实三维空间效果。三维动画是画中的景物有正面、侧面和反面，调整三维空间的视点，能够看到不同的内容。

二维动画中计算机的作用在于输入和编辑关键帧，计算和生成中间帧，定义和显示运动路径，交互给画面上色，产生特技效果，实现画面与声音同步，控制运动系列的记录等。常用二维动画软件是 Animator Studio，这是基于 Windows 系统下的一种集动画制作，图像处理，音乐编辑，音乐合成等多种功能为一体的二维动画制作软件是 Flash，交互动画制作工具，在网页制作及多媒体课程中应用；Retas 是日本生产的一种计算机二维动画软件；Pegs 是法国生产的一种计算机二维动画软件。

三维动画：制作三维动画首先要创建物体模型，然后让这些物体在空间动起来，如移动、旋转、变形、变色，再通过打灯光等生成栩栩如生的画面。创作一个三维动画的过程包括造型、动画、绘图。常用三维动画软件有 MAYA、AutoDesk、3DSMAX、TrueSpace、Lightscape、Lightwave3D、SoftImage、Animation Master 等。计算机动画使用的硬件有：SGI(Silicon Graphics Inc.)Indy ,Indigo2, PC, Macintosh, Elastic Reality, Amiga, Commodore。

6. 计算机动画的发展趋势

计算机动画是图形学的一个研究热点，涉及的相关学科比较多，比如物理学，机器人学，生物学，心理学，人工智能，多媒体技术，虚拟现实等。未来发展的方向是开发出具有人的意识的虚拟角色的动画系统，系统应具备的能力有：虚拟角色自动产生自然的行为；提高运动的复杂性和真实性：关节运动真实性，虚拟角色、收、面部等身体各部分行为的真实性；减少运动描述的复杂性；大型化、网络化、标准化。

8.3.2 The PGPGrid Project项目分析

在商业世界，不断增加的项目都在要求有一些公司联合起来负责，项目的整个生命周期要由几个公司负责，但是具体的工作又被细分开来。类似的项目有很多，比如航空宇宙工业领域，媒体领域（比如电影、电视等）。这样的协作工作在计算机科学和网格社区中被称为虚拟组织。网格除了给计算和数据资源提供了无缝访问的机会之外，还被广泛认可为工业和商业的虚拟组织提供了一个核心平台。PGPGrid 项目的目的在于探索网格在这个目的下的应用情况。该项目将三个优秀的但有很大区别的组织联合到一起，分别是：计算机动画公司派伯格斯特 Pepper's Ghost Productions Ltd.；3D 图像处理技术的专家格拉斯哥大学的 3D-Matic 实验室；欧洲领先的高性能计算中心爱丁堡大学并行计算中心（Edinburgh Parallel Computing Centre, EPCC, ）。

每个合作者在这个虚拟组织中所扮演的不同角色，具体为：派伯格斯特公司提供了本项目中的核心动画需求，使用 3D-Matic 实验室和 EPCC 的资源来制做一个短篇的计算机生成的动画电影；3D-Matic 实验室提供了对格拉斯哥大学 3D 捕获设备访问途径，从而使得派伯格斯特公司能够制作出具有很高细节程度的动画模型，为

后面的描绘做准备；EPCC 提供了中间件和访问处理器访问途径，可以访问动画模型产品和最终的绘图输出。

本项目的成熟度处于初始研究阶段，类型是核心工业型项目，获得了 DTI 的 11.25 万英镑资助资金。该项目负责人为 Paul Cockshott 博士。

PGP 网格项目目的在于将网格作为访问必要资源的方式来并行化从实验 3D 扫描仪中抽取数据的过程。这个应用项目具有潜在的高并行性，然而也具有一些特别的特点，比如实时进程的快速产生，动态交互进程网络拓扑。这些特点都需要将 Globus toolkit 中常见任务迁移能力增强。

本项目中的工作大致可以分为以下四个方面：建立基于网络的虚拟组织并且研究需求；派伯格斯特公司在 3D-Matic 实验室的动画序列捕捉；通过使用网格资源和合适的中间件，将动画序列处理为动态的卡通人物模型；在 EPCC 使用由 Pepper's Ghost 提供的模型数据和全情景信息来生成最终的场景描绘，同时也使用了网格资源和合适的中间件。

图 8-8 是在实验过程中，3D 扫描仪的数据流和工作过程。

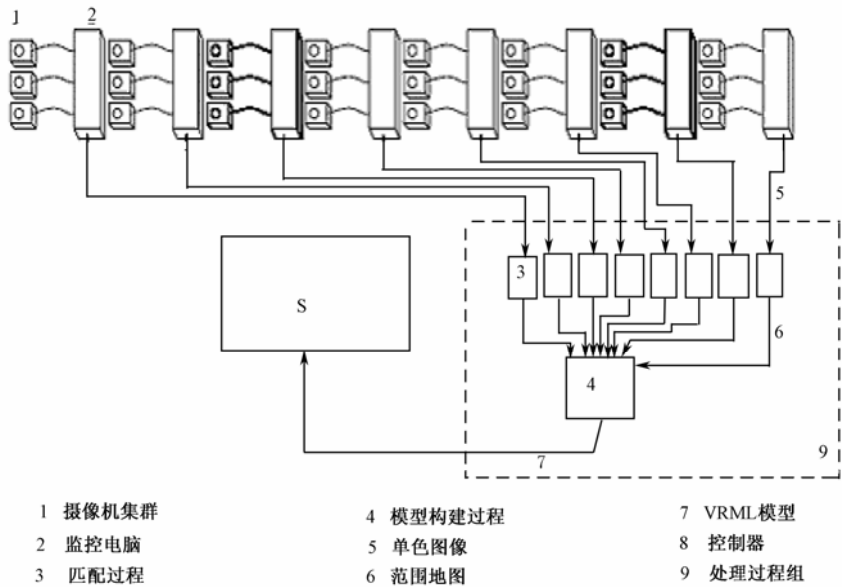


图 8-8 3D扫描仪的数据流和工作过程¹

摄像机安装成摄像机集群，每组都包含 2 个 8-bit 单色摄像机和一个单独的 16-bit 的彩色摄像机。每个摄像机集群都由一台电脑来控制，负责捕捉三台摄像机的视频并存储到内存缓冲中。下面每个同步的单色图像对都会被匹配来产生一个范围地图

¹ Viktor Yarmolenko etc., PGP Project [EB/OL]. (2008-8-20), <http://www.nesc.ac.uk/events/ahm2004/presentations/96.ppt>

(图 8-9)。来自 8 个摄像机集群的范围地图给出了点集云状数据，建模器通过分析这些数据来建造一个由三角形而成的 VRML (Virtual Reality Modeling Language, 虚拟现实建模语言) 模型。最后，连续的 VRML 模型可被组成一个空间或时间模型。其中的算法是由 Java 来实现的。

到目前为止进行过的并行实验使用了如图 8-9 所示的处理架构，其中有两个静态固定的匹配处理器还有一个简单的服务获取系统。四个摄像机集群和与他们关联的控制电脑也被使用。摄像机集群将采集到的数据以 1.2MB 每帧的速度传输给匹配进程，匹配进程匹配完以后又以 1.2MB 每帧的速度传输给模型构造器，最后生成 3.2MB 每帧的文件输出。

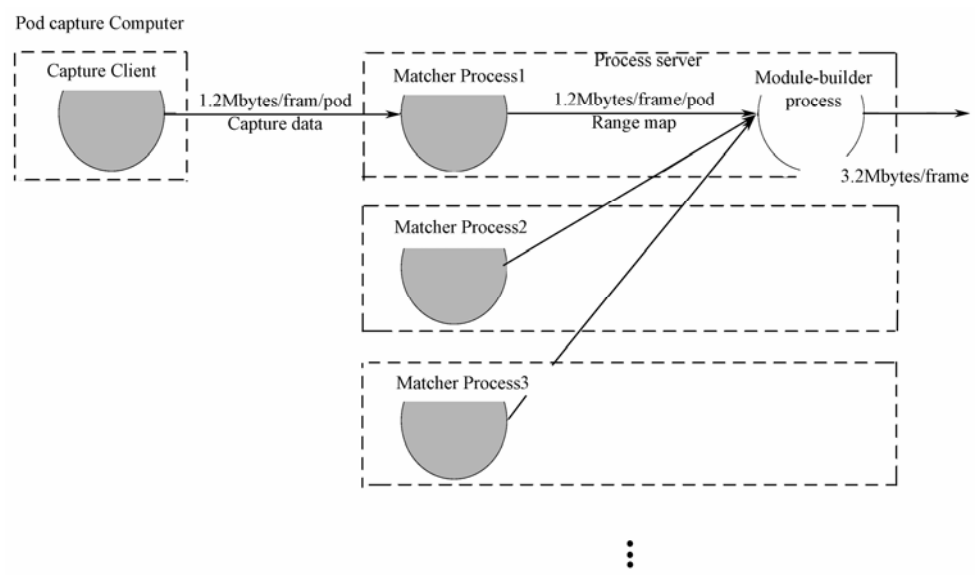


图 8-9 处理服务器的基本并行数据流 ¹

¹ Ewan Borland etc. The PGPGGrid Project , [EB/OL]. (2008-8-20), <http://www.allhands.org.uk/submissions/papers/96.pdf>

第 9 章 虚拟研究团队的组建和协同技术

e-Science 发展会碰到两个具有挑战性的问题。第一，如何满足科学家们对大规模计算和数据资源的快捷访问；第二，如何提供一种合作研究环境，来支持各种分布式的研究团队的研究活动。由于科学家们总是需要大量时间模拟他们的研究，他们对计算和数据资源有很迫切的需求。同时他们还需要能够直接同本领域的其他科学家交换多种不同类型的资源。基于以上原因，虚拟研究环境逐步发展起来了。在“英国虚拟研究环境路线图”（Roadmap for a UK Virtual Research Environment）中，网格被认为是希望提供这种面向大型数据库的广泛的计算资源和存储的架构。

9.1 虚拟组织技术

9.1.1 虚拟组织的概念

虚拟组织（Virtual Organization, VO）的发展，源于网格计算中资源共享的需求。资源共享是网格的基础。这里提到的共享并不只是简单的文件交换，而是对计算机、软件、数据等数据源进行直接访问的控制策略问题。这些问题已经在很多工业生产和科学研究的问题解决和资源代理策略中被广泛提出。准确点说，这些需求必须是可以高度控制的，资源提供者和资源使用者能够清楚地定义和精确判断哪些资源是共享的，允许谁访问，而这些共享会在什么时候什么情况下发生。由这些共享规则所定义出来的个人和机构就形成了我们所说的虚拟组织。

在网格中，无论是资源，还是资源提供者，或者资源利用者，都是分布式的。他们每个人对资源的需求不同，要求也不同，访问方式更不一样，因此，网格中存在的共享关系是千变万化的。为了适应这些变化，有必要对他们的共享要求所需要的技术需求进行仔细研究，生成一套普遍的需求。一般来说，这些需求包括高度灵活的共享策略、更复杂和准确的共享资源控制级别、更多种类的共享资源、更多变化的使用模式等。

当前的分布式计算技术并没有开始着手解决以上提到的需求和关注点。比如说，当前的因特网技术主要致力于计算机间的联系和信息交换，并没有提供一种整合的方法能够对多个计算站点的资源进行应用。企业分布计算技术，比如说 COBRA 和 Enterprise Java，只实现了在一个单独的组织内部进行资源共享。DCE（Distributed Computing Environment, 分布式计算环境）支持站点间的安全资源共享，但是它的这种模式会让大多数的虚拟组织感到很麻烦，不能灵活运用。正在兴起的“分布式计算”公司试图利用全球的闲置计算机的计算能力，但是到现在只能对这些资源进行高度集中的访问。一句话，当前的技术要么不能适应这种大量的资源共享类型的需求，要么不能提供建立 VO 所需的共享关系灵活控制。

举例来说，在某些情况下，很多互不信任同时也有着不同等级的优先权的来访者需要共享一些资源完成一些任务。当然，共享不是简单的资源交换。共享是有条件的。资源提供者都共享了自己的资源，但是给资源的访问加上了限制，比如说什么时候，谁，在哪里可以访问，都可以进行什么操作。例如，在图 9-1 中，在 VO P 中的一员可能会允许他的 VO 同伴们只能对“简单”问题激活他们的虚拟服务。而资源使用者也会对他们所需要的资源的属性加以约束。仍然以图 9-1 为例，VO Q 中的一个访问者只会接受那些被验证为“secure”的共享资源。这些约束的执行需要一种特殊策略机制来确定资源使用者或者资源提供者的身份认证,同时也要确定一个操作是否同预定的共享机制是一致的授权。

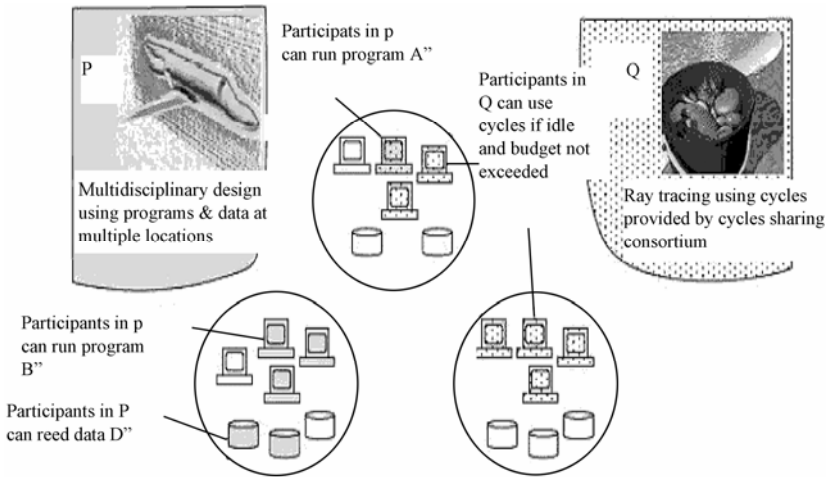


图 9-1 虚拟组织的架构¹

共享关系是会随着时间大幅度的变化的，包括相关资源的变化，被允许的访问的性质的变化和被允许访问的参与者的变化等。而且这些关系并不是一定要与一些明确的个人相关，但是有可能会被主管资源访问的策略含蓄的定义。比如说，一个虚拟组织可能会允许任何能够证明自己是“customer”或者“student”的人的访问。

这些共享策略的本质的可变性意味着我们需要一种机制来发现和定义那些会在特殊时机发生的关系的本质。比如，在图 9-1 中，一个新加入 VO 的用户必须明了那些资源可以访问，这些资源的质量如何，访问规则是什么样的。

共享关系不会总是简单的客户端——服务器关系，也有可能是点对点的。资源提供者也有可能是资源使用者，而且共享关系会在任何参与者的子集中产生。共享关系还会被用来调整对大量的不同归属的资源利用，比如说，在图 9-1 的 VO Q 当中，一个在一种共享计算资源上启动的计算过程接下来很可能会去访问别处的数据

¹ Ian Foster, Carl Kesselman, Steven Tuecke, The Anatomy of the Grid: Enabling Scalable Virtual Organizations [EB/OL]. (2008-8-20), <http://www.globus.org/alliance/publications/papers/anatomy.pdf>

或者启动一别处的个子计算过程。在这种情况下，能够以一种受限方式来标识权限的功能就相当重要，如同为多种资源互操作的协调而建立一种机制。

同样的资源也可能以不同的方式利用，这取决于为了共享和共享目的所做出的约束。比如，一台计算机在一种共享安排下能够被用来运行一段特殊的程序，然而在另外一种共享安排下它也能够提供一般的计算功能。由于缺少关于一种资源是如何被利用的先验知识，性能规格、预期和限制就会被加入到资源的共享利用策略中。

这些属性和需求勾勒出了我们所说的虚拟组织 VO，使它成为了现代计算的基础概念。VO 让不同的组织或者个人能够以一种有限的方式来共享资源，从而合作达成一个共同的目标。

9.1.2 DAME的动态虚拟组织技术

DAME（Distributed Aircraft Maintenance Environment，分布式飞行器维护环境）是最早对虚拟组织有强烈需求的项目之一（图 9-2）。这个项目采取了一种分布式结构，项目的目标是为在机场的维护队伍提供飞机引擎的详细信息，从而决定如何进行预维护。于是 DAME 系统成为了一种能够对已知问题进行分析的专家系统，同时也被作为一种决策支持系统，利用专家诊断系统来为分析那些系统没有自动识别出来的问题进行支持。

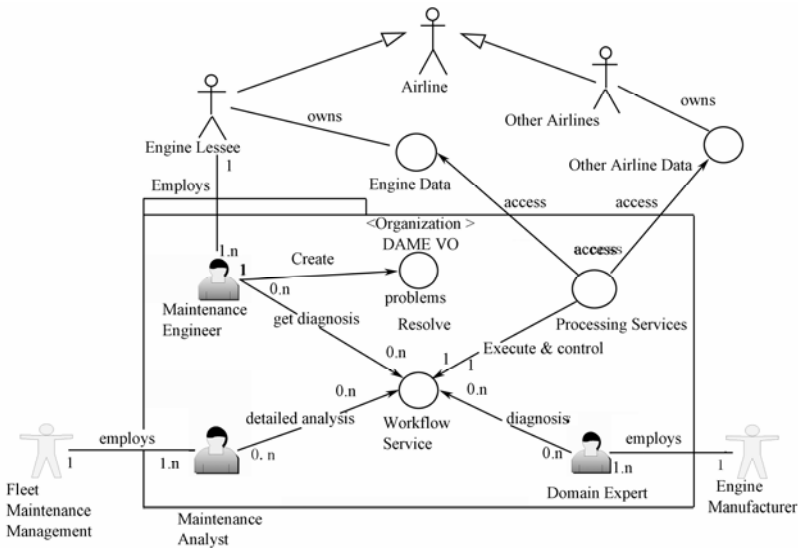


图 9-2 DAME VO¹

¹ Duncan Russell, Peter Dew & Karim Djemame, Access Control for Dynamic Virtual Organisations [EB/OL]. (2008-8-20), <http://www.allhands.org.uk/submissions/papers/140.pdf>

DAME 的目的是为引擎诊断提供一种分布式支持，因此需要大量的计算资源。图 9-2 展示的是地理上分布的人员之间进行的协作。在这个图中，人员一共被分为三种角色：维护工程师，工作在飞机场，实施飞机的维护工作；维护分析家，工作于诊断分析中心；领域专家，工作在引擎制造设计中心，资深的引擎设计专家。

DAME 的工作流程如下：当飞机降落在机场的时候，自动下载新的引擎信息，通过地面支持系统（Ground Support System）提交到网格当中。这时，DAME 就会给出一个初步检查，这个检查会返回三种信息，分别是：没有错误（No Fault Found）；检测到已知问题（Known Condition Detected）；未知状态（Unknown Condition）。对于前两种信息，维护工程师有足够的信息去解决飞机维护问题；但是如果 DAME 返回的是第三种“未知状态（Unknown Condition）”，包括飞机引擎的工作信息，就会被向上提交到维护分析家手中。维护分析家可以对 DAME 的分析过程进行复查，而且也有可能重新进行一次检查和诊断。如果维护分析家还是不能给出一个确定的分析结果，这个问题就继续向上提交到领域专家。领域专家就可以利用更多的专家工具再对这个问题进行更深层次的分析。

在这个提交的过程中，形成了一个 VO，它拥有自己分布的用户群。在这个环境中，员工们会在一种分布式环境下进行合作，他们来自不同的组，拥有不同的对工具的使用权限和对数据的访问权限。而且，在某种情况下，领域专家进行分析的时候，飞机已经起飞，而这个时候，身在机场的维护工程师已经不再属于这个 VO 了。

对这个问题，还有很多安全问题需要讨论：在某种情况下，引擎的信息是否是被拥有这台飞机的航空公司所拥有的；引擎信息是有商业价值的，应该受到保护，不能由任何不使用 DAME 或者 DAME 之外的用户来访问；基于模式匹配和案例的推理系统利用的是以前的引擎信息和诊断历史，因此需要有一种机制来控制对输出信息访问的详细程度，这取决于用户角色和用户所述的组织的性质。

9.1.3 ICENI：虚拟组织管理门户

计算与数据网格的出现带来了研究组织的变化，以动态分布式架构为中心的虚拟组织得到了充分的发展。VO 的动态本质决定了 VO 需要一种机制来管理当前用户和资源。这就需要 VOM 来实现这种功能。VOM（Virtual Organization Management）是对 VO 进行管理的系统（图 9-3），它通过 Portal 远程管理 VO，将 VO 信息上传或下载到 VOM 数据库中，还可以利用客户端，通过 GSI（Grid Security Infrastructure，网络安全架构）进行安全的认证连接，实现各种服务的交互。

使用 VOM 可以集中对网格架构进行管理，ICENI 则是 VOM 的一个典型代表。

ICENI（Imperial College e-Science Networked Infrastructure，皇家学院 e-Science 网络架构），是一种网格中间件，用户可以通过它使用一个利用分散的资源形成的图形界面建立自己的应用。在 ICENI 当中，不同的管理功能——比如说角色管理——可以通过他们自己的基于 XML 的配置文件或者通过一个图形界面来进行管理。

在这个 VOM 中，一共有三种角色，分别是普通用户、资源管理者和 VO 管理员。

普通用户就是那些想利用资源的个体，他们被归属为同一种 VO。他们如果想使用网格，首先要有一个 VO 接受的证书。加入一个 VO，进行一次合作需要两个步骤：认证（authentication）和授权（authorisation）。认证就是用户的身份证明，一般是封装在一个与 Globus 兼容的 X509 公共密钥中。而授权的获得取决于 VO 对资源的管理策略。VOM 提供了一个安全的基于 Web 的用户注册功能。用户首先用他们的网格证书向 VO 注册，这样他的认证信息就会嵌入到他的桌面浏览器信息当中。这种机制就是让用户登录后所有的与 Portal 交互的过程中都保持一种登录的状态，而是否接受或者拒绝用户对资源的访问就完全取决于 VO 的管理员了。一旦管理员准许了用户的请求，用户就可以继续他的资源访问。

资源管理者的任务是准许或者拒绝那些想要访问自己的资源的用户的请求，如果准许，它就会给用户分配一个账号。但是资源管理者也需要安装一个客户端，用来记录他的资源的使用情况，另外就是用来管理他的资源在网格上的镜像文件。

VO 管理员的任务是登记所有加入 VO 的用户和资源的信息。同时他还有给用户分配资源和监视所有资源使用情况的任务。

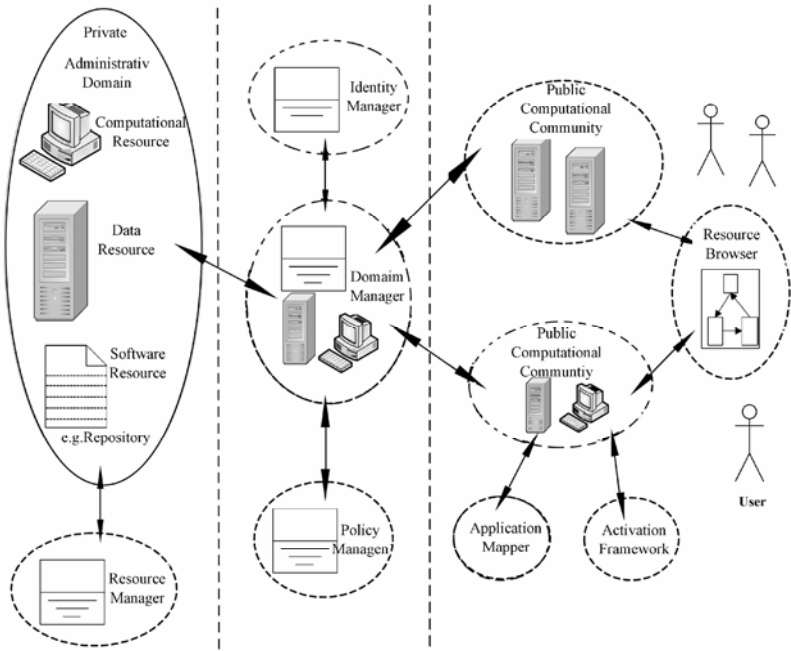


图 9-3 VOM¹

CAS（Community Authorisation Service），团体认证服务：这是一个由 Globus

¹ Asif Saleem etc, Using the VOM portal to manage policy within Globus Toolkit, Community Authorisation Service & ICENI resources,[EB/OL]. (2008-8-20), <http://www.allhands.org.uk/submissions/papers/164.pdf>

开发小组提出的授权框架，用来控制对 VO 中的资源的访问。通过它，资源管理者可以指定数据访问机制，让 VO 作为一个整体来进行控制，这样减少了 VO 管理员的工作。管理员不会经常进行注册用户、改变权限等工作。已经制订好的访问规则有足够的权限来轻松地增加新的用户，授权其访问 VO 中部分甚至是所有的资源。相比标准的 Globus 模式，CAS 具有更灵活的特点。

VO 管理员利用 CAS 来管理 VO 所信任的关系，同时它也会生成良好定义的对 VO 的资源的访问控制。当一个用户访问 CAS 服务器的资源时，他用的是 CAS 服务器所发布的证书，而不是他自己的 GSI 证书。用户首先利用自己的个人代理向服务器发送一个请求，如果权限允许，CAS 会颁发给用户一个受限的代理，其中嵌入了允许用户进行所请求的行为的授权。利用这个代理，用户就可以进行任何 Globus 所能的操作，即 GridFTP。这时所访问的资源就会应用它的本地策略，根据 CAS 的证书对本次访问进行进一步控制。这样，用户被授予的权限就和授予用户的权限与资源提供者作为一个整体授予 VO 的权限的交集相等了。

9.1.4 eMineral Project: 计算门户框架

eMinerals 项目 (the Environment from the molecular level, 分子层面的环境) 由 NERC 资助，其目的是进行有关环境问题的基础课题研究，比如核废料的保存与污染问题。此项目由巴斯大学，剑桥大学，英国中心实验室研究理事会 (Council for the Central Laboratory of the Research Council, CCLRC)，皇家理工学院和雷丁大学等单位共同进行。

eMinerals Minigrid 是一个由数据和计算资源组成的集成网格框架。除了物理资源之外，Minigrid 还包括了一套客户端工具和作为访问计算和数据服务代理的门户。

ComputePortal 是 eMinerals 开发的一个门户接口，其目的是便于让人们访问 eMinerals Minigrid 中的计算资源。ComputePortal 整合了很多其他工具，它和 Data Portal、MetaData Editor 和 MySRB 集成，形成了 eMinerals Portal。eMinerals Portal 给每个用户都提供了一个门户环境，通过这个环境可以管理所有的计算模拟和相关数据、原数据。

ComputePortal 结构上分为四个层次，自下到上分别是网格客户端接口与工具、应用层、表现层和内容管理层。底层是由客户端工具和资源库组成的，这些组件用来实现网格资源与中间件的交互。在底层之上就是表示层和应用层，依据功能的不同被分为很多不同的模块。顶层就是内容管理层，利用了 PHP Nuke 内容管理系统。使用一个内容管理系统的原因就是为了使在个性化表现模块中的碰到的意外和访问控制问题最少。

ComputePortal 在表示层和应用层都是由标准组件组成，主要模块包括：

(1) 认证 (Authentication)：这个模块的功能就是利用 PHP Nuke 框架对用户进行认证，同时获得一个有效的用户代理。另外，这个模块还会处理用户对其他门户的单独登陆。用户登录的步骤如下：用户输入用户名密码来获得证书；一旦获得了有效的证书，系统就会在本地数据库中检索出用户的详细信息，这些信息会被用于用户登录 PHP Nuke 框架，并且初始化用户的门户。

(2) 任务提交 (Job Submission)：这个模块用来将计算任务提交给 minigrid 内部的 HPC (High-Performance Computing, 高性能计算) 和 HTC (High Throughput Computing, 高产量计算) 资源，同时尽可能地将用户从等待序列中抽取出来。用户提交需求用来建立一个相应地 RSL (Resources Specification Language, 资源规范语言) 串。如果用户所请求的资源享有一个正确配置的任务管理器，那么 RSL 就会直接提交到服务器中。如果否，那么就会根据 RSL 串建立一个正确的任务提交脚本，通过 gridFTP 传输到资源层，并且通过派生出来的任务管理器进行提交。所提交的任务的 ID 号码存储在一个被任务监视模块调用的本地数据库中。

(3) 数据传输 (Data Transfer)：这个模块的作用就是将输入数据传递到资源，从资源中猎取将输出数据。输入数据可能是从客户端通过 HTTP 直接提交的，也可能是来自 SRB，或者利用搜索 DataPortal 时得到的元数据直接传输的。

(4) 资源发现 (Resource Discovery)：这个模块允许用户动态地发现、监视和查询位于 eMinerals Minigrid 内部的资源状态。它允许用户建立一个 LDAP (Lightweight Directory Access Protocol, 轻量级目录访问协议) 过滤串，并且用来查询 GIIIS (Grid Index Information Service, 网格索引信息服务) 服务器，以便找到用户所需要的资源。虽然当前项目中的各个机构的资源列表都是相对静态的，动态列表发现却是网格的一个核心概念，这个模块的作用在未来无疑会被大大加强。

(5) 任务监视 (Job Monitoring)：这个模块允许用户利用个人队列系统对他们的任务进行监视，并且可以初始化一个快速处理任务。门户将从任务提交模块获得的任务 ID 进行缓存，并且会阶段性地轮询 Globus 任务管理器，以便获得用户任务的状态。当用户重新登录到门户当中时，他就可以查看他所有的最近进行的任务状态。

(6) 元数据/日志 (MetaData/logging)：这个模块就是为所有的其他模块提供接口，来记录所有的日志和互操作，将其存入本地数据库，并且从中分析状态信息。

ComputePortal 本来应该和 eMinerals Minigrid 中的其他门户整合到一块的，这样来构成整个的 eMinerals Portal。从用户的角度来看，这将是一个无缝整合环境。

如果要将很多不同的门户整合到一块，那么最重要的一步就是如何实现他们的单一登录。这就需要利用 Web 服务和 GSI (Grid Security Infrastructure, 网格安全架构) 授权。当一个用户试图在两个门户之间进行切换，当前门户就会将当前对话 ID 放在 HTTP 头中发送出去。新的门户就会接收这个对话 ID。如果这个对话 ID 与这个门户

的对话不相匹配的时候, 用户就会向 minigrid 中其他的对话管理者进行查询, 确定这个是否和其他的门户的对话匹配。如果匹配, 那么用户的证书就会通过 Web 服务利用 GSI 认证从一个门户中传递到另一个门户中, 正如图 9-4 所示。这时门户就会在获得用户证书后继续其登录进程。

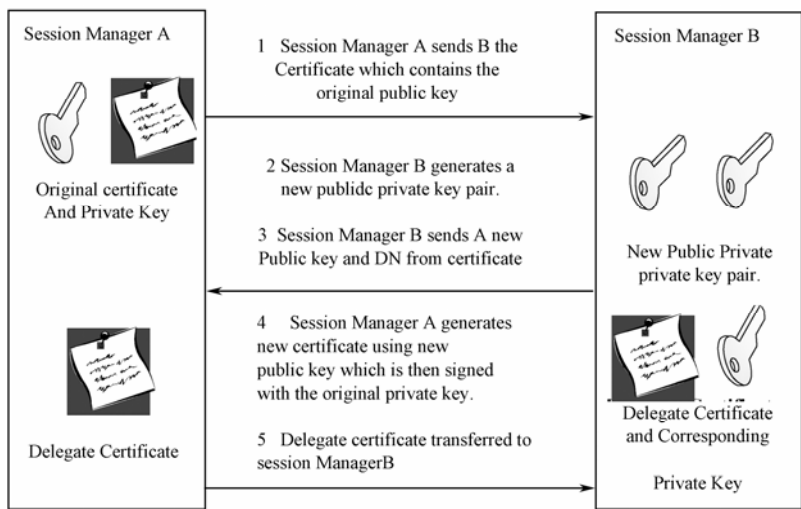


图 9-4 eMinerals¹

9.2 虚拟研究环境技术

9.2.1 概念

现在的 e-Science 的概念已经越来越宽, 逐渐进化成了 e-research。并且将社会学和艺术以及人文科学吸纳进来。同时必须指出, 不同的团体都处在对同一种技术的不同的认识时期。这样对一个大规模的国际科学合作组织的需求可要比把那些分散的人文科学家组织起来进行合作难得多。因此必须找到一种解决方案。

新的发展让研究的过程越来越复杂, 需求越来越大。VRE (Virtual Research Enviroment, 虚拟研究环境) 的目标就是利用一种特殊设计的架构支持研究队伍内部的行为, 帮助科学家们管理这种复杂性。然而难点是如何建立和支撑这种基于一个简单基础的理想的架构, 来增强科学家们的产率和效率。对那些整天思考如何建立这个架构、这个架构的潜在用户群、用户所属的机构、资金从哪里来的人们来说,

¹ R Tyer, Portal Framework for Computation within the eMinerals Project, [EB/OL]. (2008-8-20), <http://www.allhands.org.uk/2004/proceedings/papers/224.pdf>

如何满足这个需求的确是个挑战。

VRE 的一个重要指标是看它的容纳能力，而不是看它所包括组件的多少来决定其性能。因为后者会随着时间的进化而不断变量。

一个 VRE 应该有以下特性：

① 支持引导性研究过程，包括信息的组织，学术演讲的发布，合作组织的创建和维护，跨学科、机构和国家的会议和研究工作的组织。

② 应该尽可能是基于松散结构的，分布式的，可互操作的工具，而不是一个个的独立软件。

③ 应该根据用户的需求进行设计，同时兼顾可用性和可访问性，有对新工具开发的完善的评估机制。

④ 简单易用，用户只需在自己的电脑上安装一个软件就能访问，而不是要更多的专家知识。对于一些特殊领域的工具，应该集成在一起。

⑤ 采用完善的开放标准。

⑥ 安全，可信任，这样 VRE 组件就可以同那些分散在各地的有着跨机构的认证和授权机制进行交互。

⑦ 责任监控，有着足够的日志和真实性，包括对原数据的查询。

⑧ 同其他那些广泛应用的系统服务兼容，至少应该包括 Web，E-mail，实时信息，Wikis 和电视会议工具等。

⑨ 通过轻松的授权、发布、发现和访问机制来支持创建和共享资源。还包括支持适当的原数据模式，支持元数据的自动生成。

⑩ 可以进行新工具扩展，可以利用那些已经发布的标准和软件开发包、软件库等。必须尽可能容易地与现存的软件兼容，当然也包括那些私有软件。

⑪ 开源，有严格的标准约束。

⑫ 可以由用户个人或者集体自行设定它的环境，来展现出他们的兴趣和偏好。

⑬ 支持将日常任务代理给智能代理。

综上所述，VRE 应该是一套应用、服务和资源的集合，通过一些基于标准和面向服务的，在 IT 界和科学研究中很流行的框架。

9.2.2 虚拟研究环境的技术基础

面向服务架构（Service Oriented Architecture，SOA）就是一种可以将服务联合起来并提供整合能力的方法。SOA 建立在 Web 服务的基础之上，参考工业标准建立和整合了分布式的系统。图 9-5 表示了用来建立一个有着适当用户接口的 VRE 的 SOA 的一些特点。

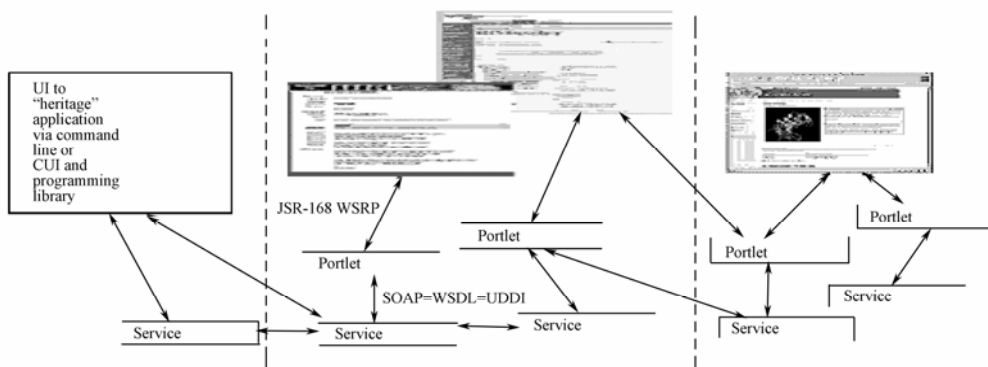


图 9-5 SOA 架构

图 9-6 表明了服务是如何被应用在这个典型的关于门户和其他客户端软件工具的四层结构图中。

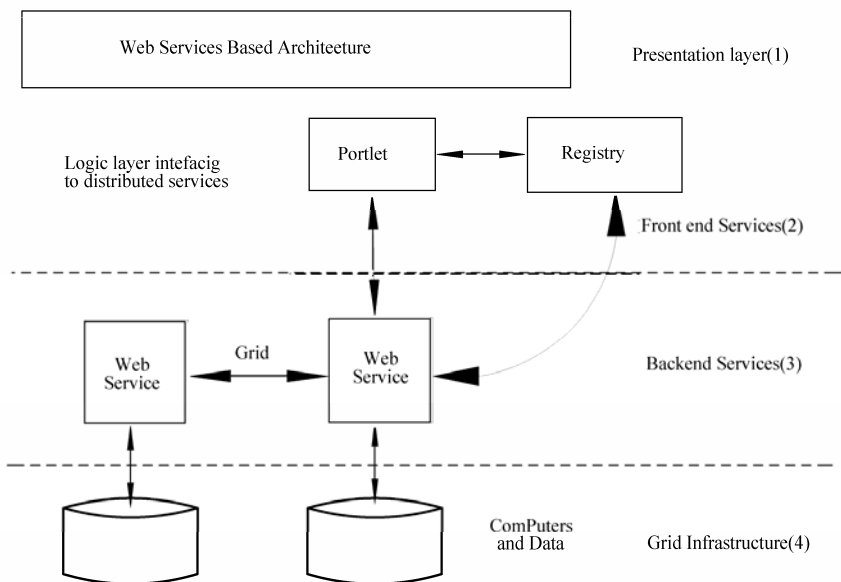


图 9-6 SOA 四层结构图

(1) 表现层，例如门户等。如果在这里使用 **sakai** 或者 **chef** 就可以给此层增加一个内容管理系统和其他的服务接口。

(2) 前端服务，逻辑层通过 **JSR-168 portlet** 应用程序接口来同他交互。

(3) 后端层，通过浏览器或者其他知名协议进行访问。可以是分布式的，也可以同其他形成一个服务网络的 **Web** 服务进行交互。

(4) 网格架构，在网格上的远程资源。

利用这种基于 Web 门户的方式，科学家们可以在任何地方，只要通过一个浏览器和因特网连接，就可以虚拟地登录这种虚拟研究环境。然而，如果使用门户这种方式，科学家们想使用现有的资源和同其他科学家进行交流，就必须提前在这个环境的统一管理处进行注册。而这种集中式的管理，并不见得就与科学研究的那种分散的、弱联系的本质相符合，而且多种团队的合作也是很有可能发生的。

从某种角度讲，点对点（p2p）也是一种分布式计算模式。如今这种分布式计算模式在很多类似于文件共享的商业运作中获得了很大的成功，例如 emule 等。它的这种计算模式可以消除现有的集中式合作环境之间的鸿沟，同时也可以将现在分散独立的科学研究有效的集中起来。点对点的最大特点就是能够突破基于 Web 门户的边界，将各个网格系统集中起来，领入一个点对点的合作环境，将网格和点对点的技术优势都发挥出来。除了计算能力和存储能力的共享外，这个架构更关注对科学家们日常合作行为的支持。首先需要建立一个网格集，来提供足够的计算资源和数据资源，而在其上就是一个点对点的合作环境。一般来说，对网格部分的管理相当复杂，也需要很强的安全保证；而这个点对点的环境在一个公共的域当中，可以由用户自行配制和管理。这个架构的设计就是为了将复杂的网格部分和点对点环境中的日常合作活动分开，这样就可以建立一种灵活的服务机制。在点对点环境下，每个科学家都会获得一份点对点应用终端。这个桌面应用其实就是一个集成研究环境的窗口，通过这个窗口科学家可以进行查询，也可以访问网格集提供的计算和数据资源。同时，这个应用也允许科学家们以一种点对点的方式进行合作，比如建立一个兴趣组，直接共享自己的工作数据和网格集所提供的信息等。

点对点环境是由一个点对点应用的网络环境组成的。每个点对点的应用都包含一个用户接口（User Interface）、一个服务客户端（Service Client）、一个服务发布系统（Service Publication）和一个发现代理（Discovery Agent）、一组团体服务（Community Services）和一组节点核心组件（Peer Core Component）。用户接口是用户访问点对点应用功能的主要通路。服务客户端主要功能是同网格提供的高级服务和注册服务双向交互。它允许利用 API 接口对这些服务进行浏览。服务发布系统和发现代理在这个合作架构中有着很重要的地位。它们提供了两个基本功能：将有关服务的信息发布到点对点环境中，发现其他同位体已经发布的服务的信息。已经发布到点对点环境中的服务的信息包括服务名、服务的 URI 及它所需要的输入数据和输出数据。已经发布的信息还包括提供这个服务的网格的一些信息。团体服务包括对团体内部日常合作的支持组件，比如说文件共享、工作组的构建和实时信息的发布等。通过这些组件科学家可以建立相关的兴趣组或者

研究社区。科学家可以向同一个组中的其他科学家直接共享自己的资源。节点核心组件包括与其他节点通信的组件、节点认证和节点资源发现组件。服务发布组件和发现代理还有团体服务就需要这些核心组件来同其他节点在点对点的环境中进行通信。

9.2.3 SAKAI: VO中间件

The SAKAI Demonstrator 项目是由 JISC 资助的。作为虚拟研究环境项目的一部分，它的目标是研究利用各种合作工具和网格服务来构建一种通用的门户，并且将这项工作的成果应用到所支持的社团建设中。

在一个门户当中，内部服务可以被用来解决框架当中所有工具的协调问题。解决方法可以以一种内部类库的形式提供，就像 Portlet 接口或者其他服务接口一样。在 SAKAI 项目中，建设了一些额外的“集成接口”，它是对原来接口的扩展，同样可以满足 e-Research 任务的特别需求。这些扩展包括以下几个部分：

(1) 描述合作研究的 XML 语法的认证与说明。这项工作就是要解决用一种标准的、便于机器处理的方式对一个合作研究对话进行描述。需要描述的部分有开始时间、过程、描述主题问题的术语等。这些文本性质的描述会被其他的工具所使用。

(2) 服务认证和身份确认系统。其目的是解决参与合作的人员的身份认证问题。几乎所有的基于 Web 服务的网格项目都采取了 GSI 或者其他的认证系统。

(3) 语义门户服务。其目的就是利用一些语义服务，让门户看起来要更“聪明”一些，可以在幕后提供一些智能推理。

(4) 基于 JSF 的 Web 服务接口生成器。在 Sakai 框架内，如果要将 Web 和网格服务作为一种工具提供访问，那么需要提供一个供用户输入参数的接口。在本项目中这种接口是利用由抽象 XML 框架描述和最终 Java 服务器型用户接口所组成的管线而成的。服务可以通过注册，比如 UDDI，而被发现。一个 VRE 用户可以检索它所需要的 Web 服务，集成到工具中，选择自己最想要的，随时随地建立自己的用户接口。

(5) 点对点服务与工具。如果要将一些流行的点对点服务和工具集成到 Sakai 框架中，首先要做的就是对项目进行一次回顾，对已存在的项目工具进行研究，找出那些会从点对点交互中获益的，并最终形成一套正确的集成策略，这套策略必须包括一个完整的 Sakai，还必须符合安全框架的要求。其次就是需要开发 JSR 168 和 WSRP 组件，一次让 Sakai 能与外部的点对点系统的交互中获益。这就是说，这个系统需要供合作的工具，比如说讨论板，实时信息，文件共享，日程表和 IRC 等。最后一步就是建立一种可以与 NaradaBrokering 和 portlet 间交互的组件。

9.2.4 IB VRE：对研究过程的支持

IB VRE (Integrative Biology Virtual Research Environment, 一体化的生物学虚拟研究环境) 项目是一个由 JISC 资助的项目, 目的是开发一个大规模的虚拟研究环境示范系统, 研究当前的合作框架, 以便对整个大规模的国际研究联盟的研究进行支持。这里提到的大规模的国际研究联盟就是 IB (Integrative Biology, 一体化生物学) e-Science 试验项目。这个项目基于复杂的环境, 对支持基于网格的各种研究工具和服务有着很高的要求。IB 的目标就是提供一种单独的集成环境, 对整个研究过程提供支持, 从试验阶段到模拟数据生成、查询、分析和恢复, 从对高性能 PC 和实验资源的访问到项目管理、项目运营、项目的教学工具支持。IB VRE 基于 WARP 和 JSR 168 兼容门户容器, 允许工具重用, 有一个分层架构。它的关键组件就是对“研究过程”的支持。“研究过程”的特点就是, 它既有过程的工作特点——由 workflow 决定, 同时还有社会特点——就是可以利用已有的工具来进行合作和项目管理。

IB 正在忙于建设两个医疗模型建设——心脏模型和癌症模型。在各个心脏模型建造者之间的合作已经逐渐进化为机构与机构之间的合作。这些合作已经持续了多年, 其中经过了联合研究过程、人员借调过程和联合论文提交过程。每个合作者都有着自已拿手的领域和兴趣, 而且双方都知道这种对双方的研究技能和兴趣都有很大影响的联合项目会很快地解决哪些关键的研究问题。

这种合作的生命周期由七部分组成: 在选定领域中识别论文和领域专家知识; 确定潜在的投资合作伙伴; 实时文本联系; 实时视频联系; 实时可视化和写作指导; 管理工作流; 信息发布管理。

IBVRE 的另外一个建设内容就是癌症建模。肿瘤的形成和生长计算建模的发展比心脏建模要差一点。现在的合作内容更倾向于勾画出这个研究的特点, 并以此为这个部分建造一个框架。癌症建模团队可以从心脏建模团队那里获得很多经验, 最终会从前边提到的那些附加能力中受益。

在 IB 项目中进行合作研究并不仅限于科学家, 事实上很多技术专家都分散在各个技术资源或者研究所中。他们可以从英国之外的各个组织中对这些科学家进行支持, 同时 VRE 环境也可以促进这种合作。比较主要的应用应该包括合作计划、出版管理、源代码管理和测试, 还有日记和日程表管理。

VRE 基于 WSRP 和 JSR-168 兼容的门户容器, 以保证所开发工具可以在任何基于 portlet 的 VRE 环境中使用。这也就允许相关的 WSRP portlet 和 JSP-168 portlet 可以在 IB VRE 中使用, 就像特别为 IB 开发的 portlet 一样。在扩展面向服务的 IB 基础架构的时候, IB VRE 采用了一种分层式或者分列式的结构框架, 包括表现层、前端

层、业务逻辑层和资源层（图 9-7）。

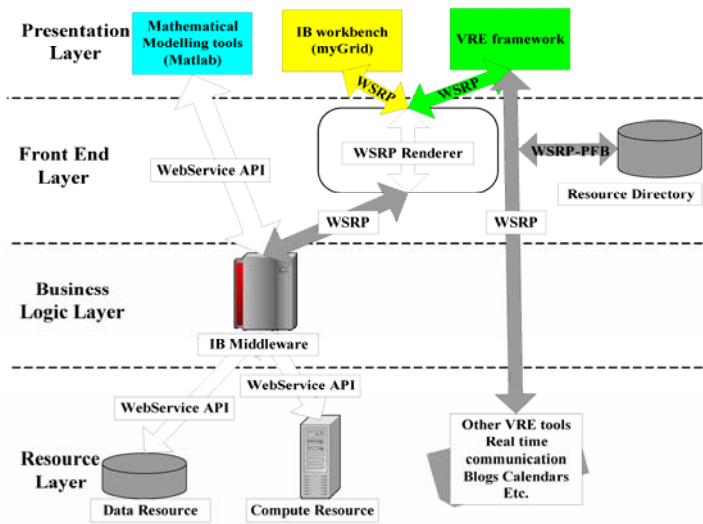


图 9-7 IB VRE¹

(1) 表现层

用户的客户端必须是轻量级的，Web 浏览器是一个理想的选择。然而，早期的 IB 项目经验却给出一些很完善的用户接口能力，尤其是复杂结果集的交互可视化，会对客户端提出额外的辅助要求。另外，定制和扩展性需求也会对客户端提出额外要求。这些额外要求绝对不是强制性的，因为用户可以通过一个 Web 浏览器来对 VRE 的核心功能进行访问和探索。在用户客户端后边应该是一个运行在 IB 架构上的门户服务器，它提供通过同等 portlet 对潜在设施和服务进行访问。

(2) 前端层

IB portlet 提供了对 IB 核心功能，例如资源定位、任务组合、任务递交等的访问途经。另外一个工作流 portlet 允许用户将这些功能按照相关的试验进行合并和管理。于是这就需要将 portlet 加入这个集合，以便提供对特殊领域的功能的大范围访问。

(3) 业务逻辑层

在 portlet 背后是潜在的服务。这些服务可以在内部被分为三个层次：合作层、功能服务层和基础服务层。基础服务层提供了对计算和数字资源的直接访问，但是这些资源一般都位于远端服务器。功能服务层会将对这些资源的访问集合起来，以

¹ Matthew Dovey, Development of a VRE to Support the Integrative Biology Research Consortium [EB/OL]. (2008-8-20), <http://www.nesc.ac.uk/talks/ahm2005/545.ppt>

一种同用户相关的并且易于理解的功能服务表示出来。大多数系统工作都发生在这一层，并且用户会通过他们的 portlet 和门户直接同这些服务进行交互。而合作层帮助用户管理这些功能服务。这一层是建立虚拟研究环境最重要的一层，将工作流程从管理任务上升到了管理用户的研究环境。在这一层里另外一个重要的组件就是解决来源信息的管理。在 IB 项目中，科学结果的来源管理是一个很重要的主题，通过它可以获得更多的相关信息，用户也可以给这些原始数据附加上科学而且有意义的元数据。

(4) 资源层

在 IB 中潜在的资源是可计算的、可建模的、以及可存储的软硬件系统。在 VRE 中，可以找到很多工具，如项目管理工具、个人信息管理工具、会议和消息管理、以及教学工具等。除了 VRE 已经集成的工具之外，VRE 也需要有一种机制，能够集成其他的 VRE 的工具。

9.3 学术交流技术

学术交流 (Sholarly Communication) 是学者与同行之间共享研究成果、公开展示自己观点的过程。通过共享，可以激发知识创新。在 e-Science 环境下，学术交流的方式发生了革命性的变化，一种全新的学术交流模式正在出现。eBank UK 项目是新型学术交流模式的代表。

eBank UK 项目启动于 2003 年 9 月 1 日，由英国 JISC 作为语义网络研究计划的一部分进行资助，由英国 Bath 大学 UKOLN (UK Office for Library Networking) 领导，英国南安普顿大学的 Combechem 项目和曼彻斯特大学的 PSIgate 物理学信息门户参与研究。该项目是在 JISC 建设 Information Environment 的背景下启动的，目标是通过化学家、数字图书馆员以及计算机学家跨领域的合作，使用如 OAI-PMH 等通用技术，实现研究数据与数字图书馆的整合。eBank-UK 项目选择结晶学 (化学的子学科) 作为试验领域，使用南安普顿大学化学院的 NCS (EPSRC National Crystallography Service) 的数据作为示例。

eBank UK 项目共包括两个阶段的研究。第一阶段 (2003-2004) 的主要工作有：在南安普顿大学通过使用经过修改的 eprints 软件建立了一个 OAI-PMH 存储库；大量结晶学数据被存放到存储库中，并提供一个针对结晶学定制的本地浏览和检索界面；基于都柏林核心集设计了一套元数据框架，以描述数据；实现了元数据收割，并建立了数据发布的范例；开发用户场景以帮助理解处理复杂数字对象的需求。第二阶段 (2005 至今) 的主要工作有：复杂数字对象描述研究，及其在描述科学数据中的应用；将研究数据连接到 e-learning 中，开发其在教育方面的

应用等。

e-Bank UK 选用 OAI-PMH 体系结构。系统包括 e-data 和 e-print 两个存储库，其中 e-data 存储库是由南安普顿大学开发的存放结晶学研究数据的存储库，在存储结晶学试验数据的过程中，数据的描述信息一部分是人工录入，一部分是自动生成的，e-print 存储库是由 ePrints UK 项目开发的 OAI 论文存储库。中央服务通过 OAI-PMH 收割存储库中的元数据，并将研究数据与研究出版物的元数据一起进行索引，其中“中央服务”即 OAI-PMH 中所谓的服务提供者，同时也是 e-Bank UK 项目中所谓的数据汇总。当 eBank UK 数据汇总收集完元数据后，便会产生一个 SGML (Standard Generalized Markup Language, 标准通用置标语言)索引，并使用 Cheshire 搜索引擎向用户提供检索功能，用户能够通过一个检索界面检索数据及相关文献。其中的元数据包括一个指向原数据集的链接，使用户可以访问原始数据。从不同存储库中收割的元数据不仅提供了一个对异构数据统一的访问入口，而且提供诸如使用来源与出版物的主题词检索数据集等的增强型功能。另外，Cheshire 还提供机器接口，支持诸如基于 SOAP 的 Search and RetrieveWebService (SRW) 协议和 Z39.50 协议等，为将搜索服务嵌入到其他框架中提供了基础，也正是基于这一点，再加上使用基于 CGI 的机制和信息门户有关的标准，e-Bank UK 将检索服务嵌入到了 PSIGate 信息门户中。

在 e-Bank UK 整个的处理过程中，数据流情况如图 9-8 所示。

知识是嵌入在元数据中的，所提供服务的水平直接取决于元数据的质量、数量及内容。eBank UK 在原有结晶学试验数据集元数据的基础上选择元素，并通过重新设计的 DC schema 进行呈现，以发布给数据汇总和信息门户。具体元素及其类型和描述如表 9-1 所示。

另外，为描述复杂数字对象，eBank UK 使用 METS (Metadata Encoding and Transmission Standard) 对元数据进行封装，以传输核心元素以外的元素。

为实现数据与研究出版物的互操作，一方面出版物和研究数据都需要标识和定位，目前已有很多标准和机制，如各种 URI 和 OpenURL 等可供选择；另一方面两者在数据结构和元数据上必须存在某些一致性，并需要对这些共性进行引导连接，eBank UK 的解决方案是使用出版物的参考文献信息连接研究数据，但实际情况是，这部分数据并不都是可以开放获取的，而支持 OAI 访问的就更少，所以 eBank UK 目前只能使用一些出版商提供的样本元数据进行试验。

续表

Data Name	Data Description	Data Type
Publication_date	Date of releasing ePrint to eBank/world	String
Last_revised_date	Date ePrint last revised	String
Keywords	Categories	String (set?)
Scheme	2D diagram	String
ICHI	International Chemical Identifier	String

9.4 协作工具

对支持协作的信息工具的开发已经有很长的历史,从上世纪 80 年代中期的 CSCW (Computer Supported Cooperative Work, 计算机支持和协同工作) 开始,人们就在对远程协作的实现方式和动态协作有关问题的理解方面取得了重要的进步。CSCW 是一种独立的技术,其发展并不是建立在特定需求的基础上,而是依靠对社会、对人们交流和协作活动的观察,试图通过开发相关的技术来辅助交流协作的过程,其目标是通过使用计算机技术在时间和地点维度上的支持进一步加强组织工作的进程。在计算机技术飞速发展的今天,对于 CSCW 系统的实现也相应发生着巨大的变化,从技术思想的进步,到应用程序和工具的实现,再到各种协议标准的提出,都在推动着人类的交流协作活动;另一方面,随着社会生产力的发展,人类生活水平的提高,以及技术对于人们需求的拉动,社会对于交流协作也在不断提出更高的要求。具体到 e-Science 这样的环境背景下,首先大科学的前提对科研交流协作提出了强烈的需求,而虚拟组织、虚拟实验室等概念的提出则表现了交流协作观念的更新,基于网格的各种先进技术与项目研究成果又为新世纪的交流协作提供了强有力的技术支持。下面就对 e-Science 环境下的交流协作技术作进一步介绍。

9.4.1 基于网格的协作工具——Access Grid

传统的交流协作模式建立在正式的但并不频繁的面对面的会议上,Email 等非正式的联系方式主要用于传递帮助或建议等信息以及各种文件资料,另外合作者间还阅读对方发送的手册和文档资料。总体来说,这就是传统合作者所期待的交流协作。但是,大规模的科研合作所提出的交流协作要求并不仅仅局限于此,例如,对于视频会议、程序共享的需求等。也许对于其中部分需求,已有的方法和工具似乎已经提出很好的解决方案,如通过 IM (Instant Message) 也可以实现视频会议,但在面临高级交流协作需求的时候,这些工具和方法却又显得无能为力。一个简单的例子是,尽管使用 MSN 或微软网络会议等软件可以实现视频音频传送的功能,但是却无法很好地实现视频会

议中多方同时参与的功能，更不要说多人同步操作、共享应用程序等功能了。

1. 基于网络的交流协作概述

从分布计算发展起来的网格技术，为支持交流协作的应用和系统提供了新的框架和模式。其实，从某种意义上说，网格中所有组件都是为了进行协作而设计和使用的。但是，其中大部分是用来进行协同工作的，如协同计算、协同数据存储/管理/访问、协同监测等等，只有少数的组件是为交流协作所使用的，具体说来，基于网络的交流协作实际是指使用了网格的基础设施和安全认证服务等组件，实现的跨网格的组织间交互，如大规模的分布式会议、协作工作对话、研讨会、培训与讲座等，其优点如下所列，典型的实例如 Access Grid。其优点如下所述：

- 安全的协作环境。
- 稳定的数据传输（利用网格的基础设施实现音频/视频流的低丢包率，从而提供稳定的交流协作服务）。
- 健壮并可扩展的系统框架。
- 与其他网格服务形成的整合服务体系。

2. Access Grid的特点

Access Grid 是一组由美国 Argonne National Laboratory 管理的用于支持 Internet 上全球范围内组织间交流的网格系统。目前，拥有上百个访问节点，遍布 47 个国家，其中包括中国地区 9 个节点，并已经发放了数千个用户证书。通过 Access Grid 的访问节点，用户可以与世界范围内的其他拥有访问节点的机构进行视频会议。

尽管基于网络的交流协作系统具有很多的优点，可以解决复杂的问题，但并不是说传统的交流协作工具就要被取代，相反地，需要根据实际的需要和能力有选择地使用交流协作工具或者结合使用。常用的交流工具包括：eMail；Instant Messaging（IM）；Webtools：网站、wiki；Access Grid。

如果从获取成本、安装成本和交流时间延迟三个角度对这四种工具进行评估的话，结果如图 9-9 所示。其中，获取成本（Access Cost）表示每次使用的困难程度；安装成本（Setup Cost）表示建立交流系统的困难程度；交流时间延迟（Time Delay）表示交互的时间延迟。上述所有色条越短，越能表明在该工具在该项指标上表现良好。由图 9-9 可以看出，E-mail 的建设费用最低，而且使用方便，但交流时延较长；即时消息使用也很方便，但建设和时延情况一般；基于网络的交流协作系统的建设成本是最高的，每次使用起来也是最麻烦的，但交流时延是最短的；而网站或 wiki 在各方面情况都不理想。

因此，在实际部署协作环境的时候，需要综合考虑各方因素，协调使用多种协作工具，最终达到服务科研，方便交流的目的。

Access Grid 不同于传统的电话或因特网交流方式，其特点体现在如下五个方面：

- （1）可扩展的虚拟会场服务。
- （2）安全控制。
- （3）应用程序共享和动态工作区连接。

- (4) 节点管理及用户界面。
- (5) 异步协作能力。

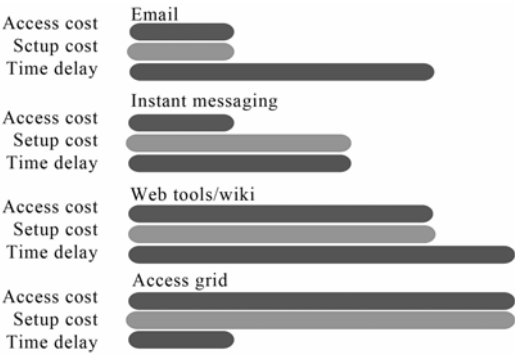


图 9-9 主要交流工具的粗略比较

基于网络的交流协作系统体系结构如图 9-10 所示。

其中需要指出的是，Access Grid 不同于普通的基于单机的交流工具，但也并不完全属于网格计算技术，而只是使用了网格的基础设施和安全认证服务，但也正因为如此，为网格中间件提供了一个可视化环境的接口，使得人们可以将 Access Grid 整合到 e-Science 当中，为组织间交流提供有力的协作支持，一个有力的证据是，所有英国的 e-Science 中心都拥有 Access Grid 节点，并为大量研究人员所使用。

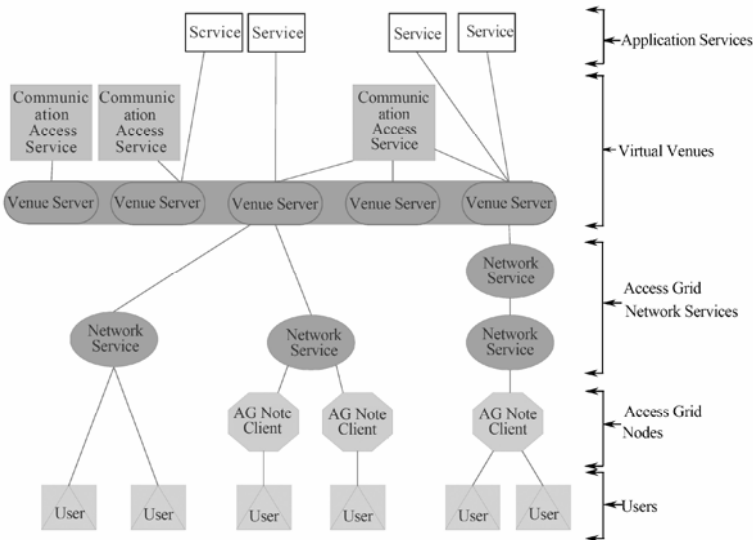


图 9-10 基于网络的交流协作系统体系结构

9.4.2 虚拟组织建设——eMinerals

eMinerals (Environment from the Molecular Level) 是 NERC e-Science 测试平台计划中的一个项目, 其目标是将原子模拟工具的发展与网格计算技术相结合以在不断趋近现实的条件下, 实现在环境问题上的模拟研究。

1. eMminerals的研究内容

项目主要关注四个方面的研究:

(1) 科学应用方面, 关注核废料封装、土壤粒子对有机污染物的吸收, 以及矿物的风化和沉淀等问题。

(2) 应用程序代码开发方面, 关注相关科学应用的应用工具代码的开发, 如 DL_POLY3、SIESTA、Casino 等的开发。

(3) 网格技术方面, 关注支持研究的网格基础设施的开发、建立, 具体包括 eMinerals minigrid、SRB、eMinerals 数据门户及元数据插入工具、eMinerals Condor 计算池。

(4) 在虚拟组织建设领域, 关注协作工具的开发和使用, 包括 MAST (英文全称是什么)、Personal Access Grid、及相关机制的研究。

其中, 在虚拟组织建设方面, 由于项目包括来自剑桥大学、巴斯大学、雷丁大学、伯克贝克学院、伦敦学院、英国皇家学院和 CCLRC 实验室的研究队伍, 形成了三个小组负责管理项目。这三个小组分别是网格开发小组 (grid team), 模拟代码和方法的开发小组 (application team), 以及科学应用小组 (Science team)。项目的成员研究机构与这三个小组之间需要进行交流协作, 这是 eMinerals 项目成功的关键。

2. eMinerals中虚拟组织的描述

为了最大化提升地理上分布的参与者之间的交流数量, 项目采用了虚拟组织的概念和思想, 具体实现描述如下:

(1) 在安全认证方面, 由于 Minigrid 使用 Globus 组件进行安全封装, 所以 eMinerals 的各合作机构在进行安全认证的时候都使用 X.509 协议。

(2) 在标准协议方面, 所有合作机构均使用 CML (Chemical Markup Language) 作为描述模拟数据的标准以实现互操作。

(3) 在网格层次上, 使用网格标准 (WSRF) 实现互操作, 由专门人员与 Condor 团队合作开发 WSRF-enabled Condor, 并使用 Minigrid 作为测试平台。

(4) 在协作工具方面, 除了传统的 Email、IM (Instant Message)、外部网站 www.emineral.org、内部 wiki 外, 还在 Access Grid 基础上, 打包实现了可安装在个人 PC 机上的 Personal Access Grid, 从而实现了便捷的视频会议交流方式。

另外, 项目还开发了 MAST (Multicast Application Sharing Tool) 实现了应用程序的共享, 使得在一台计算机上运行的程序界面可以被多个用户看到; 使用 OTRS (Open Ticket Resource System, 开源 email 支持工具) 实现参考咨询功能; 使用

9.4.3 e-Science协作调动空间——CoAKTinG

CoAKTinG (Collaborative Advanced Knowledge Technologies in the Grid, 网格高级合作知识技术) 项目启动于 2002 年, 历时 2 年, 由英国 e-Science 计划资助。其目标是通过先进的知识技术和应用程序推动先进的、分布式的 e-Science 协作, 诸如: 通过本体加强多种方式、多种媒体、时间同步的组织间讨论和问题求解; 基于知识的规划和任务支持以加强问题的处理和讨论; 提供学术演讲或辩论以加强协作; 提供临场感和可视化, 以加强远程协作。

另外, 该项目还特别关注整合和示范第三方机构开发的工具和技术的应用, 以支持 e-Science 协作, 诸如对 Access Grid 和 VRVS (Virtual Rooms Videoconferencing System, 虚拟房间视频会议系统) 等工具的使用。

作为项目的一个重要部分, CoAKTinG 工具在一些场景下被部署并强化, 包括 e-Response、Combechem 和火星探索。

由于 CoAKTinG 的重要影响, 两个后续项目得以启动: Co-OPR (Collaborative Operations for Personnel Recovery) 和 Memetic (Meeting Memory Technologies Informing Collaboration)。

在具体的交流协作工具方面, 该项目开发并整合了四个工具: BuddySpace: 即时消息与出席通知 (instant messaging and presence notification); Compendium: 图形会议与组织记忆记录 (graphical meeting and group memory capture); Process Panels: 智能操作列表 (intelligent “to-do” lists); 会议记录与重现 (meeting capture and replay)。

1. BuddySpace

即时消息客户端软件 (基于 Jabber 协议), 具有在线提示功能。引入人物图形可视化和图像地图呈现功能。允许多个协作工作组即时浏览并自然地给用户一个虚拟组织的整体印象。

BuddySpace 无疑为会议提供了支持, 比如可以传输文件或进行不受干扰的交流。对于像 Access Grid 这样的分布式会议, 对个体的呈现则更会给人协作的感觉。另外, BuddySpace 还可用来进行会议控制, 比如进行演讲排队和投票选举等活动。如果要记录会议, 还可以使用对话日志功能。

2. Compendium

最初作为辅助跨功能 BPR (Business Process Redesign) 团队的方法开发于 1993 年, 已经被在多个工程和研究领域项目所应用。起初创建目的是为了在团队成员间共享理解, 早期主要的功能是一个 IBIS (Issue-based Information System) 系统概念映射工具的结合, 支持非正式的探索式交流与沟通, 使用结构化的模型方法。这几

年来，经过测试和精化，又添加了更多的模型方法。

在本项目中，**Compendium** 导入知识技术设计空间不同的元素（图 9-11），使得会议成为一个真正的针对组织知识创造的事件，一个实时生产的结构化的、共同拥有的、可检索的组织记忆。作为会议的产品 **Compendium** 通过实时高质地记录工具以及会议中中介的、有效的工具抓住了知识建设的瓶颈问题。**Compendium** 是开源的核心组件，允许科学家记录其关于某个试验及结果的演讲，是试验操作与论文内部的核心连接，读者可以籍此更好地理解作者。

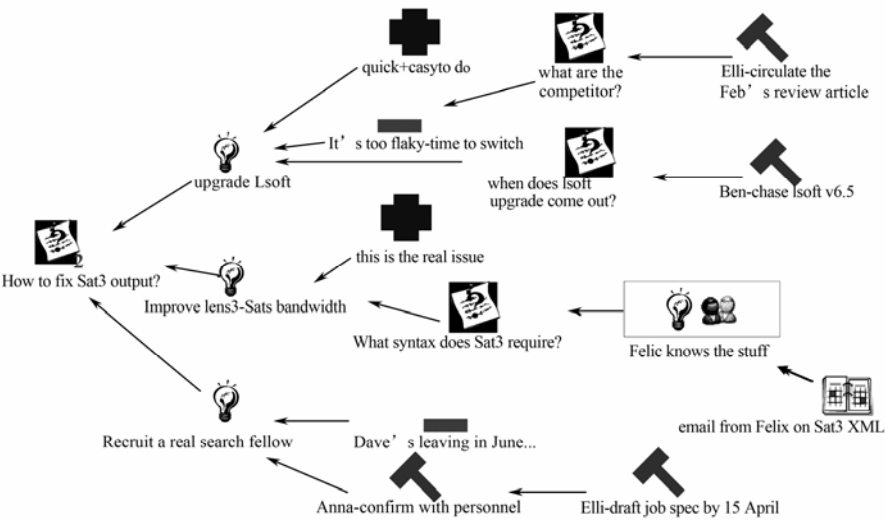


图 9-11 Compendium 纲要图

3. I-X Process Panels

I-X 是一系列工具，功能为辅助创建和修改一个或多个“产品”（如一篇文档）的过程。I-X P2 简单的界面扮演了一个智能的“to-do”列表的角色。面板显示了用户现在的问题和动作，其中使用标准操作规程管理复杂和运行时间较长的进程。I-X 还拥有个协作元素，使得问题和动作可被在不同进程面板间传送，以实现组织内的工作流。Webservices 可以被自动调用以启动相关的进程。I-X 背后的模型是基于 <I-N-C-A>约束模板。在会议场景下，会议动作被映射到<I-N-C-A>动作上，在 I-X 面板上被创建的会议动作被传送到相关的用户面板上。

4. Meeting Replay

在传统的 CoAKTinG 场景下（如 Access Grid 结点），全部的音频、视频日志都可记录，但过于冗长无法实际应用。人们希望能够从较高的层次引用会议，再“zoom-in”查看其中的细节。每一个会议都用 RDF 描述，遵守 OWL 的会议本体，诸如会议时间、地点、参会者、音频/视频记录、讲稿、以及辩论评注等。通过 RDF 描述的有子事件的事件被映射成一个传统的时间线，自动的用 HTML 和 JavaScript 发布在网站上。用户可以通过

视频时间线导航会议，或跳转，或者创建 **Compendium** 结点。通过使用共享的 **AKT** 引用本体，还可以链接到其他知识库中的资源。另外，还在时间线上增加辅助性注释。

另外，在项目中，通过在不同的工具间共享和理解本体，实现了互操作并降低了提供可用和有用结构的工作负担。

其中，为记录会议中的信息，使用本体描述以下信息：会议本身的信息；与会者信息；会议所讨论的内容信息；会议相关文件的信息，包括多媒体文件，而由于人们希望能够按照时间对会议进行索引和导航，因此还需要有表现时间的本体。在这里引入“事件”的概念，一个事件可以各种时间的描述信息，也可以分出子事件，借此建立了会议的时间结构。

CoAKTinG 会议本体扩展了 **AKT** (**Advanced Knowledge Technologies**，高级知识技术) 项目的 **OWL** 会议本体，以更好的描述会议，具体内容和关系如图 9-12 所示。

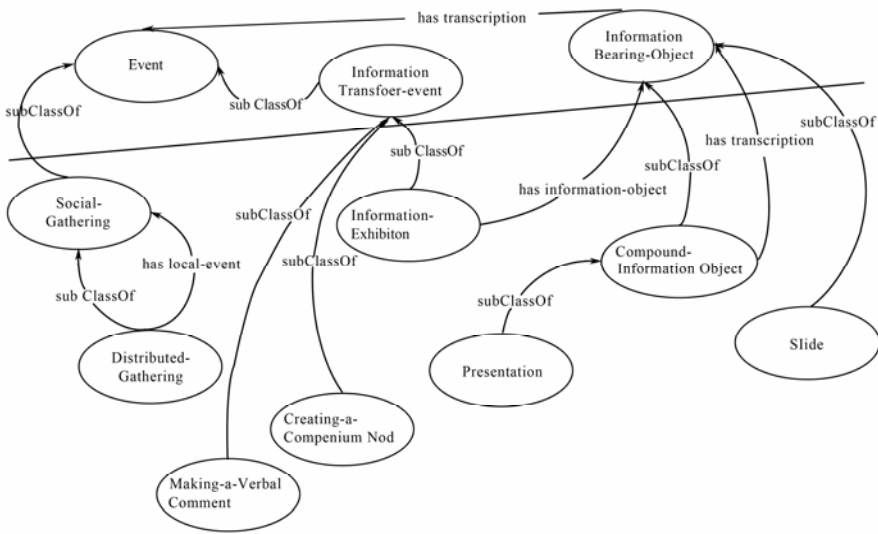


图 9-12 CoAKTinG会议本体的简单呈现¹

9.5 问题求解环境

在《基于网络的科学问题求解环境研究》一文中²，作者指出问题求解环境是人们探索如何高效利用计算机解决实际问题所取得的成果之一。用户可以在只掌握少量计算机操作知识的情况下利用它顺利地解决某领域的问题。科学问题求解环境专针对科学计算领

¹ CoAKTinG Update, [EB/OL]. (2008-8-20), http://www.actors.org/coacting/resources/presentations/CoAKTinG_Jan2005.ppt

² 王仁重,徐 铸,尹华祥.等.基于网络的科学问题求解环境研究[J].计算机工程.2006(8):99-100.

域的问题求解环境。科学计算作为计算机应用最重要的场合之一，有着其特殊之处，比如计算规模大、精度要求高、程序设计复杂、人机交互频繁等等。良好的科学问题求解环境不仅应该提供高性能的计算，还应该提供大量高效的计算库资源、易用的程序设计环境等等。而将传统的科学问题求解环境与网络结合起来，充分挖掘利用网络在资源共享和协同工作方面带来的可能性，建立起一个基于网络的科学问题求解环境系统，具有重要意义。

下面就从英国 e-Science 计划的两个具体项目出发，分析介绍基于网络的问题求解环境的建立和部署。

9.5.1 协作医学问题求解——MIAKT

MIAKT (Medical Imaging and Advanced Knowledge Technologies, 医学图像高级知识技术) 项目是由EPSRC的两个跨学科研究协作，AKT和MIAS (Medical Images and Signals, 医学图像与信号) 合作启动，于 2005 年 3 月结束。项目的目标是要将AKT和MIAS的能力应用于乳腺癌的检测和诊断领域的协作医学问题求解上，关注对专家会诊乳腺癌及规划医疗途径的协作诊断的支持。项目涉及如下研究机构：南安普顿大学 IAM 研究组；伦敦皇家学院计算图像科学研究组；英国开放大学；牛津大学医学观察实验室；谢菲尔德大学资源语言处理研究组以及 Epistemics有限公司。

MIAKT 应用建立在一个分布的框架上（如图 9-13 所示），使用基于网络的服务向通用用户提供离散的、功能全异的服务。该框架是抽象自任何应用领域的，可以提供强大的结构来迅速在一个新的领域建立一个新的知识管理应用的原型。从这个角度讲，MIAKT 只是该框架的一个具体应用。

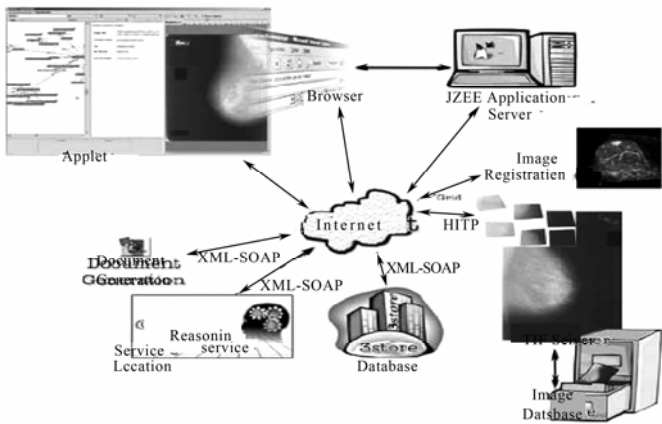


图 9-13 MIAKT框架及应用¹

¹ Nigel Shadbolt, MIAKT: Combining Grid and Web Services for Collaborative Medical Decision Making [EB/OL]. (2008-8-20), <http://www.allhands.org.uk/submissions/papers/258.pdf>

该问题求解环境共有三个不同的执行位置。服务器：将应用程序以 Java 程序的方式传递给客户端，服务器是一个 J2EE 企业版服务器，并呈现给客户端特定应用访问分布资源的透明的框架。客户端：用户使用 Java 程序运行浏览器和程序接口，这些 Java 程序基于客户端本体进行初始化；Web 资源：一系列远程的模块，增强框架的基本知识管理能力。这些 Webservices 通过应用本体被连接到服务器端的通用任务调用框架中。

1. 服务器端

服务器端向客户端提供其感兴趣的功能。除了分布式的设计外，还有应用本体，提供了向不同领域通用用户呈现不同功能的方式。服务器端框架包括四个组成部分，如图 9-14 所示。

- (1) 认证系统：基于用户数据库。
- (2) 状态整理：提供服务器上分布功能的接口。
- (3) 本体小程序：提供访问本体的服务，如检索和显示本体等。
- (4) 任务调度子系统：对分布 Web 资源进行调度。

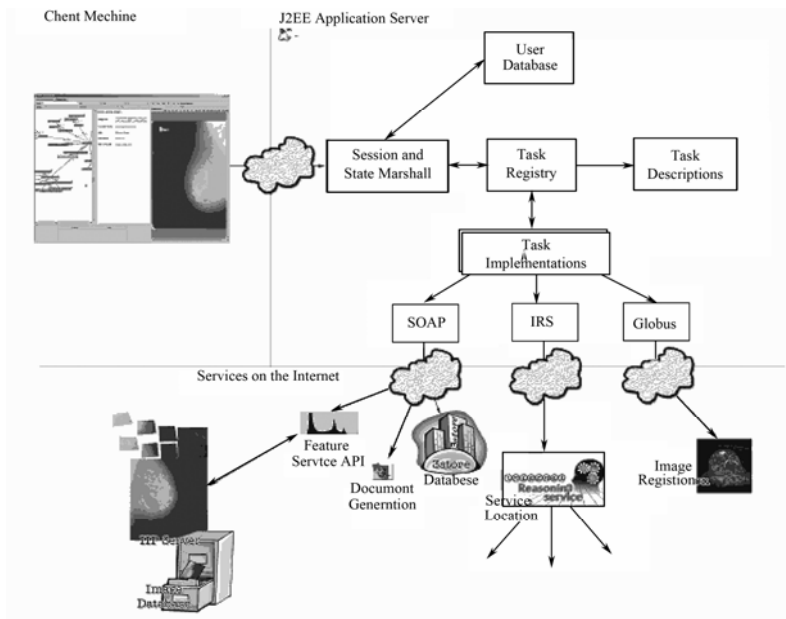
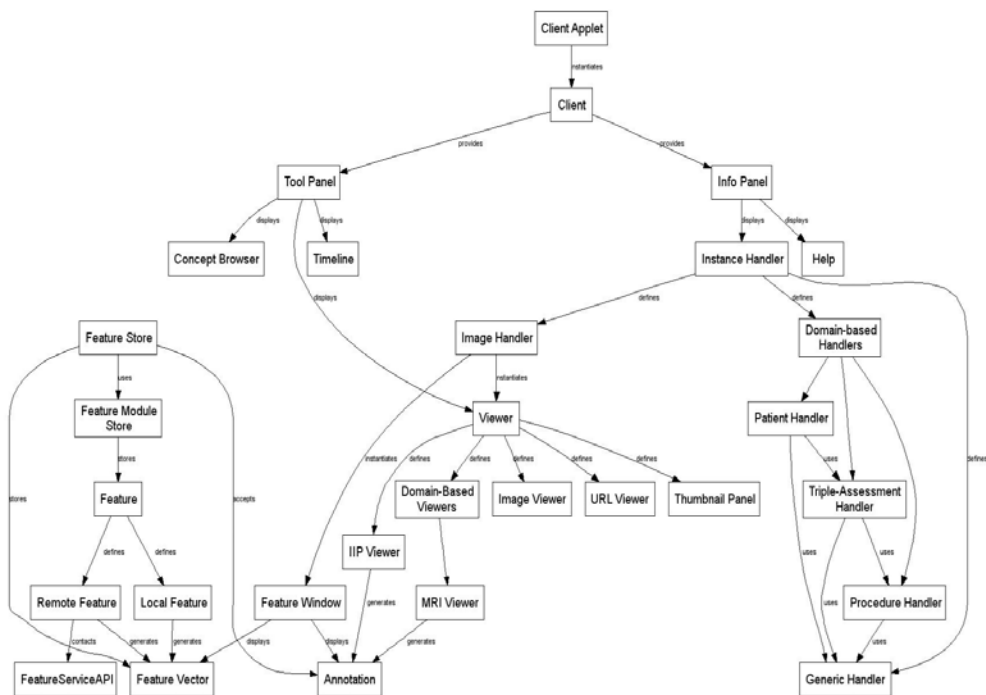


图 9-14 服务器端框架¹

¹ Nigel Shadbolt, MIAKT: Combining Grid and Web Services for Collaborative Medical Decision Making [EB/OL]. (2008-8-20), <http://www.allhands.org.uk/submissions/papers/258.pdf>

2. 客户端

客户端应用程序是从 J2EE 服务器传送的应用程序，使用 J2EE 服务器服务以提供功能，整体框架如图 9-15 所示。具体组件概述如下。

图 9-15 客户端框架¹

- (1) 浏览器：图像浏览器、URL 浏览器、缩略图面板、IIP 浏览器、MRI 图像浏览器。
- (2) 实例处理程序：用于显示本体库中特定实例的信息。
- (3) 功能子系统：处理和比较功能向量。
- (4) NativeFeature API：提供了整合由 C 或 C++ 编译的图像算法到 Java 功能框架中的方法。
- (5) 其他：搜索功能。

3. Web资源

- (1) 系统主要使用如下 Web 资源: 3Store RDF 数据库; 功能服务 Web-Service API;

¹ The MIAKT Architecture –Technical Manual,, [EB/OL]. (2008-8-20), <http://www.aktors.org/miakt/docs/The MIAKT Architecture 2005-04-08.doc>

IIP 图像服务器；IRS（Internet Reasoning Service，Internet 推理服务）。

（2）MIAKT 特定的 Web 资源：MRI 服务器；自然语言生成 Web-Service。

9.5.2 分布式飞机维护环境——DAME

DAME（Distributed Aircraft Maintenance Environment）项目，即分布式飞机维护环境，于 2002 年 1 月由英国 EPSRC 的 e-Science 计划投入 350 多万英镑启动，于 2004 年 12 月结束，为期 3 年，是 e-Science 计划的试点项目，旨在示范使用网格和 Webservices 技术（基于 OGSA）推动将地理上分布的资源 and 数据整合到一个决策支持系统，以进行诊断和维护应用系统的设计和开发，项目选用维护民用客机引擎的应用领域进行通用框架部署。

1. DAME 的具体目标

- （1）开发一个分布式的诊断网格测试平台。
- （2）在 Rolls-Royce 公司飞机引擎维护场景中示范网格测试平台的优点。
- （3）设计并建立一个利用网格中间件、关注网格中数据管理的分布诊断支持的系统框架。
- （4）研究并进一步理解数据网格的性能问题。
- （5）为非结构化、无索引的数据设计一个分布式数据存储。
- （6）开发一个网格可以使用的错误识别和诊断方法。

2. DAME 核心技术

- （1）AURA：Advanced Uncertain reasoning architecture for Pattern Matching，基于模式匹配的高级不确定推理架构。
- （2）QUOTE：基于神经网络的应用程序实时监控技术。
- （3）CBR：Case based Reasoning systems for intelligent decision support，基于实例的智能决策支持推理系统。
- （4）Webservices 框架：基于 Globus 的开放框架协议。

3. DAME 解决的在线决策支持系统问题

- （1）对远程资源的访问（如专家、计算资源、知识库等）。
- （2）系统中核心人员与参与者间的交流问题。
- （3）对信息流和数据质量的控制。
- （4）在决策支持系统中对来自全球的不同资源的数据的整合。

4. DAME提供的核心服务

(1) DAME 引擎数据服务

包含 QUATE 系统、传感器和辅助的地面连接。引擎数据服务控制引擎监测系统（QUATE）和地面站的交互，地面站建立了网格数据存储器。因为飞机到处停，所以要在多个地点提供该服务，模拟数据通过下载程序传递给白玫瑰网格的数据存储器（目前由 MySQL 实现，将转换为 Oracle）。

(2) 数据挖掘服务

使用了 AURA 相关矩阵技术与快速模式匹配，从而实现对引擎健康监测数据的检索；二层框架：AURA 通用服务层；具体应用程序代码层，直接与用户打交道。

(3) 引擎模拟服务

提供从飞行数据和运行模型中提取参数能力，目标是推断当前的引擎状态。目前 Rolls Royce 公司的引擎性能模型已经被作为一项网格服务放到网格上，基于 GT3，该引擎模拟网格服务可以被其他虚拟组织或应用程序使用。

(4) 基于事件的推理服务

整合 CBR（Case Based Reasoning，基于事件推理）服务以进行诊断过程的决策支持。CBR 系统精炼知识库并记录 DAME 的错误诊断方法。可用来管理诊断操作的工作流；维护 DAME 知识库。

(5) 数据分析服务

由领域专家部署数据分析服务，以追踪造成异常事件发生的根本原因。目前使用的数据分析工具是基于 QUOTE 系统的。另外，系统还提供可视化工具以交互评估错误数据。基于 DAME 门户提供的协作工具，诊断过程中不同的参与者可以共享处理结果，并允许远程专家共同分析错误。

参与 DAME 项目的研究机构包括：约克大学计算机学院，利兹大学计算机学院和机械工程学院，牛津大学工程学院，谢菲尔德大学自动化控制与系统工程学院，Roll-Royce 公司，Data Systems & Solutions 公司，CYBULA 有限公司等。

该项目建立起一个重要的大都市网格基础设施——白玫瑰网格（WRCG，White Rose Computational Grid）。目前，利兹大学、谢菲尔德大学和约克大学已经就此开展合作，并斥资 280 万英镑，共同致力于这一网格基础设施的建立之中。该网格建设主要是为了在专家和软件工具分布于网格的情况下，解决实时智能功能抽取、智能数据挖掘和决策支持技术。

DAME 项目的示范工程是基于一个诊断工作台环境的概念进行部署的，该工作台以门户的形式实现，对分布于白玫瑰网格的资源进行整合并提供访问，形成一个虚拟组织。整体的框架如图 9-16 所示。

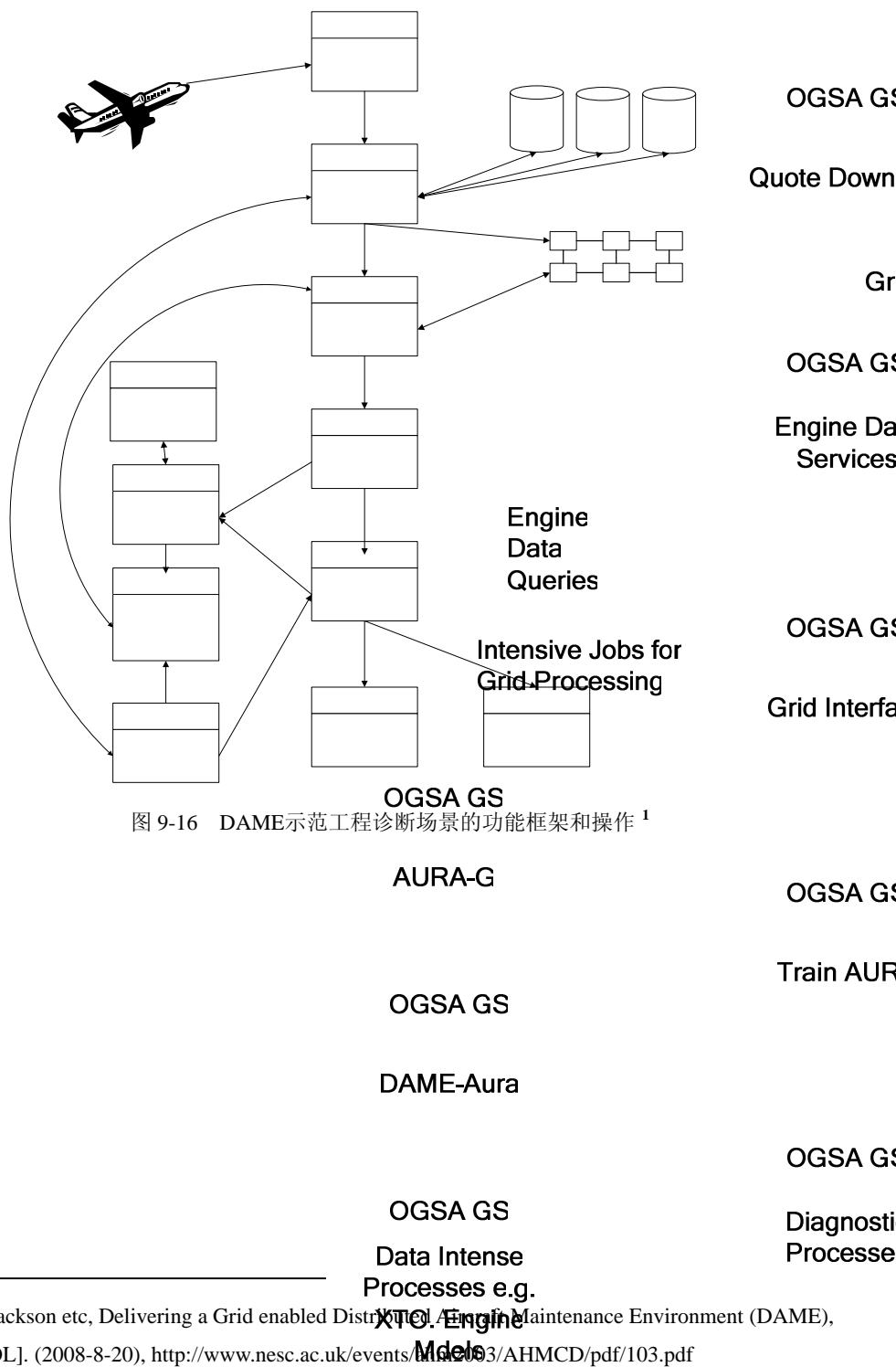


图 9-16 DAME示范工程诊断场景的功能框架和操作¹

¹ Tom Jackson etc, Delivering a Grid enabled Distributed Aircraft Maintenance Environment (DAME), [EB/OL]. (2008-8-20), <http://www.nesc.ac.uk/events/mc2003/AHMCD/pdf/103.pdf>

第三篇 规划发展篇

“交流”与“共享”是科学研究永恒的主题。网络革命性地改变了科学研究交流与共享的方式，将交流、合作与共享推广到科学研究活动本身乃至贯穿于整个科学研究过程之中，标志着科学研究共享时代的全面到来——科学家群体共享的对象不仅有传统层面的数据、资料、信息，更增加了科学家的智慧与劳动以及科学仪器设备的共享。

科学研究对象已不再是简单的孤立系统，而是涵盖更大的范围、横跨多个学科。无论是微观粒子、宏观宇宙、染色体、基因，还是航天工程、全球气候、生态环境等问题，都已经令人类的研究达到一个更为复杂深奥的程度。基于网格的科研工具和虚拟协同为科学问题的空前复杂化提供了解决方案。例如，仿真和模拟已经是复杂科学问题必不可少的研究手段。对于一些高难度、大规模的研究，由于缺乏相关的资源或条件，依靠传统的科学研究手段往往会受到诸多制约。如需要有大规模军队参与才能完成的项目，或者是需要在一些人类根本就无法进入的环境中进行实验，包括超高温、超低温环境。

在网格技术的强大支撑下，在变革科学研究模式，提高科学研究的合作、交流与共享程度，提高科学研究效率的强烈愿望的推动下，人们构造出一种全新的科研协作模式和大科学工程，这就是 **e-Science**。它以新一代互联网技术和网格技术为基础，通过 **Internet** 联合组成一个共同的虚拟研究团队，共享资源和成果，协同工作共同完成大型的现代科学研究。

e-Science 就是使用计算和通讯技术，通过全球性的合作来促进科学研究活动的开展。**e-Science** 整合了网络计算、程序设计工具、数据和可视化技术，这些使得科学家可以通过直接的仿真、实验和观察的比较，获得有关其实验研究的更深入见解，通过 **e-Science** 研究人员将可以访问到支持关系、概念和推论规则的知识库，从而为其获得新的见解提供便利。

在 **e-Science** 环境下，科研模式将发生极大的变革。与以往的科研模式不同，科学家直接面对的将不再是各种分散的数据操作，与此恰恰相反，通过网格技术及其相应的中间件实现程序，以往只能分散性进行的各种操作将被得到集成。于是，科学家只需要提交任务请求，便可以通过单一的入口、无须考虑具体实现过程地接受集成化服务，大大提高了科研效率。

在调研当前国内外 **e-Science** 发展现状的基础上，结合对 **e-Science** 关键技术深入、系统的剖析，本篇将在前两篇的基础上，探究当前国内外 **e-Science** 项目、实践在规划与管理上的规律性特征，并以中国 **e-Science** 当前 **e-Science** 发展中存在的问题为切入点，开创性地提出全面促进我国 **e-Science** 发展的策略建议。

第 10 章 e-Science 的规划与管理

10.1 e-Science 已经成为发达国家科研模式创新的方向

通过调研可以发现，各国 e-Science 规划的目标、参与规划建设国家数量及分布、启动规划建设时间以及规划周期等指标都明确表明，e-Science 已经成为世界各国科研模式创新的主流方向和首要举措，正如 20 世纪 90 年代全球各国普遍兴起“信息高速公路”规划建设，目前 e-Science 作为“下一代网络基础设施”正在成为各国带动本国科研模式创新，实现大科学梦想的首要选择。

1. 科研模式创新成为各国 e-Science 规划的重要目标

目前，包括英国在内的大多数国家与地区均希望通过实施 e-Science 计划，加强 e-Science 建设，实现科学研究的共享与协同，进而实现科研模式的转型与创新。例如，英国政府将 e-Science 核心计划的目标定位为，让来自世界各国的科学家能够以一种前所未有的高效率工作方式，协同式地开展科学研究，并在此过程中实现数据、设备、仪器、知识等多种资源的共享。在欧洲，欧盟框架计划建设 e-Science 的目标旨在使整个欧洲地区的研究者更好地分享知识和各种资源，使网络真正支撑欧洲 ERA，网络将会整合整个欧洲的世界一流的科技资本以增强欧洲的竞争力和生活质量。因此网络将是实现欧盟 2000 年设定的“里斯本战略”（即是到 2010 年使欧盟发展为世界上最有竞争力和活力的知识经济体）目标的关键技术。

2. 发达国家普遍启动 e-Science 规划

课题组调研的 679 个 e-Science 项目，由 28 个国家发起（未包括参与国），其中主持 10 个以上研究项目的国家有七个，分别是英国、美国、法国、中国、加拿大、德国和澳大利亚（图 10-1），主要分布于亚洲、欧洲、美洲和大洋洲（图 10-2）。统计表明，英国乃至整个欧洲的 e-Science 研究和建设规模最大，美国也不甘示弱，拥有大量的相关研究项目，中国在调研项目中共有 30 个，排在第四位；在按洲统计中，欧洲以 503 个项目占四个洲的第一位；美洲排在第二位，共有 122 个项目；亚洲、大洋洲等其他地区在 e-Science 的研究方面则稍显逊色。

3. 2010 年以前是发展 e-Science 的关键时期

从世界各国 e-Science 规划启动时间与建设周期看，2010 年以前是发展 e-Science 的关键时期。统计表明，尽管早在 1983 年，美国为了构建一个全球气候模型，就启

动了与 e-Science 密切相关的项目——Community, Climate, System, Model。而早在 1999 年 12 月德国已经完成重点围绕“网格中间件”的研究项目——UNICORE (Uniform Interface to Computing Resources)。但是,对 267 个项目进行综合分析(图 10-3)表明,发达国家集中在 2001—2004 年启动 e-Science 规划,2002 年是 e-Science 规划启动的高峰期,共启动 108 个项目。从项目建设周期看,各国 e-Science 建设周期多在 2~4 年(图 10-4)。其中,持续 4 年的项目数量最多,共有 96 个,占 14.7%;持续 3 年的项目 82 个,占 12.5%;持续 2 年的项目 57 个,占 8.7%。

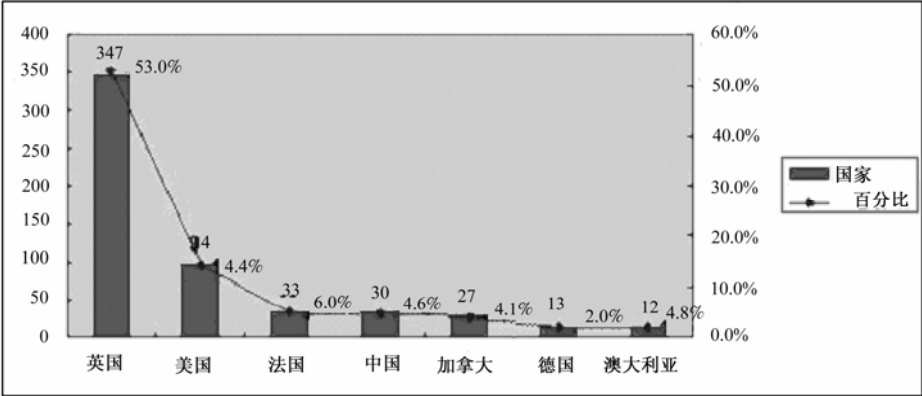


图 10-1 调研项目国家统计分析

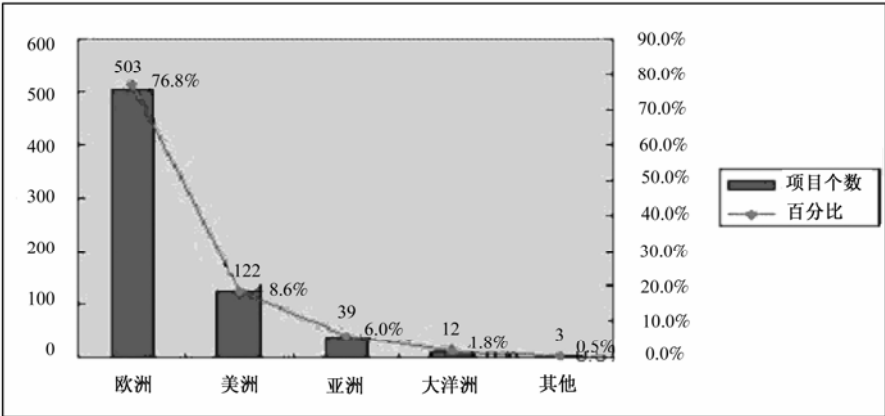


图 10-2 调研项目洲统计分析

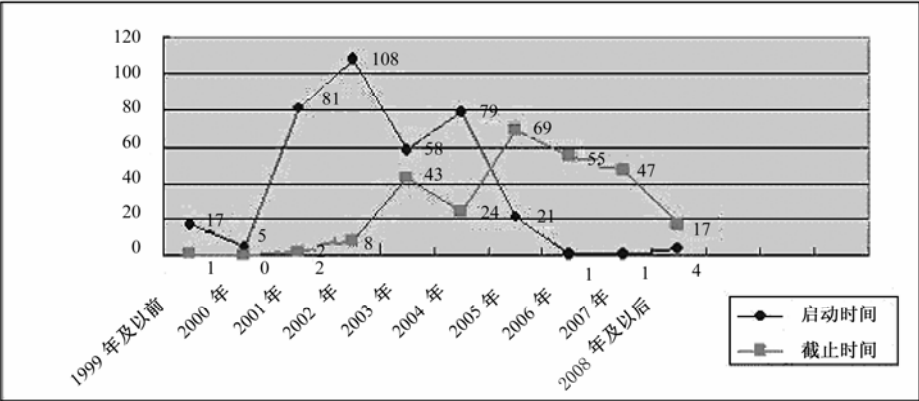


图 10-3 调研项目启动与截止时间统计分析

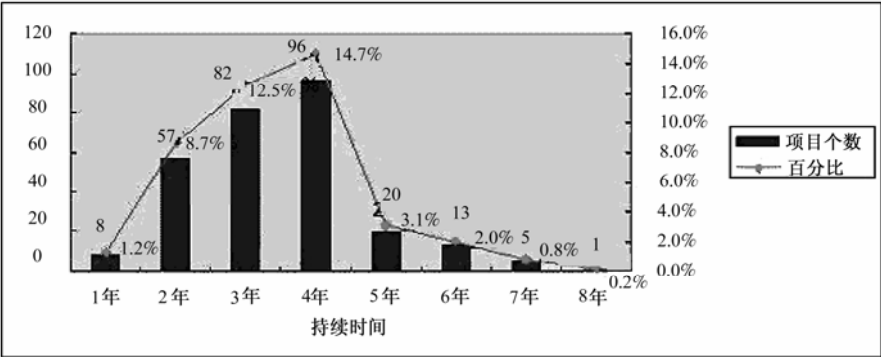


图 10-4 调研项目持续时间统计分析

10.2 网络技术成为e-Science核心技术

网络并不等同于 e-Science，却是 e-Science 的核心技术。英国研究理事总会主任 John Taylor 给 e-Science 下的定义是：“e-Science 是指关键科学领域中的全球化合作以及使这种全球化合作得以实现的下一代基础设施。”而使这场科学革命得以实现的下一代基础设施通常是指网络。使用网络这个术语来描述 e-Science 的中间件基础设施是出于对未来科研的这样一种构想：在这个基础设施中，计算资源、计算周期（computing cycles）和存储、以及昂贵的科学仪器设备和软件能够像今天的电网设备一样按需获取。

e-Science 的本质目标是：实现分布在不同地方的仪器设备、科学数据、信息资源、计算资源、科研工具等各种科学研究资源的共享，实现虚拟科研团队科学研究活动

的协同。网络是在网络的基础上通过对各种资源的集成和管理，实现对资源共享与协同工作的支持。因此，网络成为e-Science驱动各类资源共享与科研协同的重要组成部分和网络基础设施。在e-Science的旗帜下，全世界的科学家和计算机科学家共同致力于开发出一套部署在传统物理研究网络之上的软件工具与服务，这套核心中间件服务在英国被称为网络中间件，在欧洲被称为e-Infrastructure，在美国被称为Cyberinfrastructure¹。网络使科学家们有能力配置一个安全而可控的环境以共享其科学研究所需要的分布式资源。

1. 网络实质地推动e-Science发展

英国国家 e-Science 中心将 e-Science 研究与建设内容细分为 31 个小类。在此基础上，根据各小类的研究重点将其归并为 3 个大类，即网络基础设施、网络工具包与网络应用，并以此组织、管理自身承担的 231 个 e-Science 项目。在欧洲，网络研究贯穿于欧盟的 e-Science 建设过程中，在欧盟的第五框架计划中总共启动了 24 个网络相关的项目，而目前的第六框架计划也于 2004 年启动了 12 个网络项目。

(1) 网络支持了 e-Science 协同与共享模式的实现

利用网络技术，人们便可利用当前互联网架构，构建 e-Science 科研模式。在此模式下，地理上广泛分布的各种资源，包括计算资源、存储资源、带宽资源、软件资源、数据资源、信息资源、知识资源等将被整合成一个逻辑整体，并为科学人员或者是为了实现某一共同科研目标而临时组织起来的、具有动态性的虚拟组织提供一体化的信息和计算、存储、访问等应用服务，并最终使得科学人员和虚拟组织能够在一个虚拟环境下共享多种资源、开展协同工作。诚如网络技术开拓者伊安·福斯特所说，e-Science 这种全新的科研协作模式的到来，将使人们看到了科研协作的一种渐进式的发展前景，甚至可以说是一种革命性的发展前景。

(2) 网络支持了 e-Science 的各种应用与功能拓展

网络是 e-Science 得以实现的核心技术，在 e-Science 五层框架中位于具体应用层之下，直接支持着 e-Science 各种应用的实现，并拓展了 e-Science 的各种功能。例如，e-Science 具体应用层的一个重要功能是协同科研。网络中间件技术使科学家们的协同科研合作成为可能，网络将使世界各地的科学家聚集在同一个中间件之下，使科学家们使用同一种经由网络中间件翻译的“语言”操纵互联在一起的各种资源和科学仪器设备，从而实现协同工作。又如，e-Science 应用中的模拟仿真需要大量高性能的计算资源，而分布式的计算网格则使模拟仿真所需的计算资源需求得到满足。因此，网络是 e-Science 各种应用和功能扩展的核心支撑。

¹ 由于网络、e-Infrastructure 和 Cyberinfrastructure 三个术语在内涵上较为相近，只是不同地区在名称上的差异，因此，本书统一使用网络代替后两个术语。

2. e-Science激发了网格的创新

从 20 世纪 90 年代早期开始, 网格发展经历了三个阶段: ① 第一代网络的主要目的是将一些地理上分布的超级计算中心连接起来, 共享这些超级计算中心的计算资源。② 第二代网络采用具有标准化接口的中间件来处理规模和异构带来的问题, 可以支持需要进行大规模计算和海量数据处理的高性能应用程序。这个时期的核心研究项目有 Globus 和 Legion 等等。主要集中在解决通用平台所面临的异构性, 扩展性、适应性问题。人们通常将侧重提供大规模计算能力的网格称为计算网格, 而将具有海量数据处理能力的网格称作数据网格。③ 第三代网格的研究采用面向服务的方法, 具有自主计算的特性。此时的网格技术集中在分布式全球协作、以面向服务的方法为核心、信息层处理等。

(1) 从数据网格到信息网格

目前 Internet 上不乏数据, 但这些数据或者缺乏格式, 如纯文本、图片、影音文件以及大量的 HTML 页面, 或者只能依照特定格式来完成操作, 如 FTP 等。只有通过浏览 (WWW 方式) 才能知道信息是否适用, 从而使这些数据无法更好地使用, 可以说这些数据仅仅是数据, 而非信息。计算和数据网格主要解决数据访问的问题。在此层次中, 人们用网格技术将世界上不同地域、不同接口的各种设备、资源连接起来, 形成一种方便访问的途径。数据网格包括海量数据的仓储体系和网格结点之间的数据传输两大方面。信息网格建立在数据网格之上, 在原始数据和信息用户之间建立了中间“信息层”, 让异构的信息能够同构地访问, 从而使“数据”成为“信息”。

(2) 从信息网格到知识网格

得益于共享协作的科研模式, e-Science 环境下的科学发展将越来越快, 由此引致的后果就是科学分支不断增多, 交叉学科、边缘学科、综合学科层出不穷, 学科之间的关联度进一步加大。在此背景下, 科研人员已经不再满足于通过网络而获取的一般性服务 (如检索信息、收发邮件、发布信息), 而迫切希望可以通过网格对知识信息进行分析、归纳、综合, 让网格为自身研究提供知识解决方案。

事实上, 正是 e-Science 知识需求的拉动, 网格才会由数据网格进一步发展到信息网格, 并最终知识网格服务于科研人员。具体地说, 知识网格是建立在信息网格之上的一个智能互联环境, 利用数据挖掘、信息融合等技术, 在数据网格完成了数据访问途径、信息网格集成大量关联信息之后, 利用计算机挖掘隐藏在其中的“规律”, 能使用户或虚拟角色有效地获取、发布、共享和管理知识资源, 并为用户和其他服务提供所需要的知识服务, 辅助实现知识创新、协同工作、问题解决和决策支持。信息网格主要包括信息表示和检索以及 Internet 上的信息分布与组织, 而知识网格包括数据挖掘、知识挖掘、问题求解, 它将超越现有的信息检索、过滤、挖掘、问题回答等技术领域。

10.3 基本形成统一的e-Science技术体系

1. e-Science五层技术体系

综观英国、美国和欧盟等主要发达国家和组织的 e-Science 建设现状，各国在 e-Science 建设过程中基本形成了统一的技术体系。这个技术体系可以用一个五层结构来描述，自上而下分别是具体应用层、应用开发环境与工具层、网络中间件层、网络基础设施层和资源层（详见图 10-5）。



图 10-5 e-Science 五层技术体系

(1) 具体应用层

具体应用层直接面向科学家用户，满足各种科研应用中对海量计算能力、远程数据获取和昂贵科学仪器设备的共享的需求。这些具体应用包括参数模拟和重大科研难题的解决等。应用层能为科学家用户提供一个应用服务的门户，在这个门户上提交数据、访问远程资源和收集研究结果。

(2) 应用开发环境与工具层

应用开发环境与工具层为 e-Science 开发人员和用户提供各种高水平的服务，该层主要为开发人员提供开发应用程序以及测试性能与可靠性的服务，为科学家用户提供有效而方便的方式利用这些应用的服务。

(3) 网络中间件层

网络中间件层提供一般的网格应用服务和特定应用服务，这些核心服务包括远程处理管理、资源联合配置、存储访问、信息登记、安全、数据访问与转换、服务

质量（QoS）（诸如资源预定与交易等）。

（4）网格基础设施层

网格基础设施层是支持 e-Science 各种应用的基础，主要是指高带宽、连通性好的通信网络，该层主要定义资源层各种数据交换的网络通信协议。

（5）资源层

资源层是整个 e-Science 框架的最底层，这一层由各种可获得的资源构成，包括超级计算机、计算机集群、存储网络、专业数据库服务、昂贵的科学仪器设备以及可视化设备。

2. 典型项目技术框架分析

DAME（A Distributed Diagnostics Environment for Maintenance）是英国一个典型的 e-Science 项目，采用四层结构。在该模式中，最底层为网格基础设施，负责调用网格中各种可用的资源；处于最底层之上的是网格中间件层，主要负责网格服务的管理，需要提供包括资源代理、数据代理、任务调度以及服务注册在内的各种服务，是 DAME 系统层与网格基础设施的交流媒介。自最底层往上位处第三层的是 DAME 系统层，它通过利用各种网格工具软件与网格中间件进行联系，并建立 DAME 诊断门户系统将工作流支持服务、决策支持服务以及资源支持服务以集成的方式提供给用户使用。在第四层，也即是处于该模式最顶层的“维护空间”，这是用户进行维护操作以及开展商务活动的操作层。在这一层中，用户通过诊断门户系统对各种用于维护的个案数据进行操作。

3. e-Science技术研究的热点分布

在技术层面，根据英国国家 e-Science 中心的网格框架（截止于 2007 年 1 月），对调研项目进行分类，该框架共分三层，包括 31 个组件，具体项目分布情况如表 10-1 所示，在应用服务、网格计算服务和数据管理服务方面汇聚了较多的项目。

表 10-1 调研 e-Science 项目技术研究分布

所 属 层 次	小 类 名 称	项目总数（未去重）
具体应用层（301）	应用服务	183
	网格计算服务	116
	领域本体、任务本体与元数据	45
	发布与归档服务	5
	问题求解环境	39
	合作工具包	57
	数据管理服务	115
	数据挖掘与分析服务	68
	网格管理服务	71

续表

所属层次	小类名称	项目总数（未去重）
应用开发环境与工具层（287）	仪器管理服务	9
	市场经济服务	8
	模式与程序原型	18
	时序安排、监测与诊断服务	13
	虚拟与沉浸环境服务	36
	可视化服务	58
	workflows服务	34
网格层（253）	审核	6
	认证与授权	34
	代理	10
	合作与远程仪器服务	56
	协同调度	18
	数据访问服务	59
	数据编目与溯源	26
	故障管理	6
	全局事件服务	9
	网格信息服务	70
	全局请求	7
	监测	23
	服务质量的可靠性	14
	资源访问服务	47
	安全与隐私	27

10.4 大规模的合作成为各国e-Science建设的主要方式

10.4.1 跨国合作是各国e-Science建设的特征之一

1. e-Science建设展现了国际合作趋势

在本课题统计的所有 e-Science 项目中，非合作项目有 118 个，仅占全部项目的 18%，即大多数的项目都是合作型项目，在合作项目中跨国合作成为 e-Science 建设项目合作的新趋势之一。

除去国家与洲、国际组织或者其他机构之间的合作项目外，从目前统计分析看，国家与国家之间的合作主要体现的是小范围合作，合作规模一般都不太大。情况如下（如图 10-6 所示）：

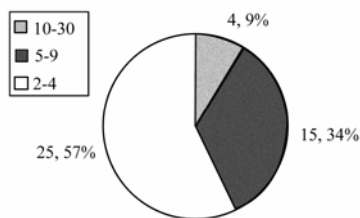


图 10-6 调研项目国家合作统计分析

有 10~30 个国家合作的项目有 4 个，占 4%；5~9 个国家合作的项目有 15 个，占 34%；2~4 个国家合作的项目有 25 个，占 57%。其中，合作度最大的项目是由美国发起的 Access Grid，一共有 32 个国家参与，分别是阿尔及利亚、澳大利亚、奥地利、巴西、加拿大、中国、克罗地亚、捷克、芬兰、德国、印度、爱尔兰、意大利、日本、韩国、荷兰、新西兰、挪威、菲律宾、波兰、葡萄牙、波多黎各、俄罗斯、新加坡、斯洛伐克、西班牙、瑞典、瑞士、泰国、英国、美国、委内瑞拉。该项目主要是集合大型资源，提供包括大量的多媒体显示、表达和互动环境，网格中间件的界面和可视化环境通过网格支持组对组的互动。目前，该项目已经拥有 47 个国家 3400 个用户的认证。

2. 德国、英国、美国、意大利、法国是国际合作的主要国家

参与合作项目 3 个以上的国家有 29 个（详见附录 G），按参与项目数目排在前 5 位的依次是：德国、英国、美国、意大利和法国。德国共参与了 30 个项目；英国参与了 28 个项目；美国参与了 21 个项目。中国参与了 9 个项目，与奥地利和希腊并列排在第 10 位（详见图 10-7）。

对 e-Science 项目国际合作的可视化聚类分析表明，处于国际合作核心的国家由英国、德国、法国、意大利、荷兰等（见图 10-8）。图中点代表国家，点的分布代表了国家之间的合作分布，箭头的指向代表国家之间的合作关系。

由图 10-8 可以看出：

（1）国家之间的合作情况仍然很大程度上取决于地理环境因素，如欧洲之间的聚类、亚洲之间的聚类等。

（2）美洲和大洋洲涉及的国家比较少，但是从图中看出，其与欧洲和亚洲之间的合作非常频繁。

（3）德国和英国成为欧洲国家聚类的核心。

e-Science 目标是实现全球范围内的科研协作，同时通过以上的统计数据，可以很明显地看到，在 e-Science 的发展过程中，各国非常重视项目的跨地域、跨行业合作。

在英国，其核心计划六大构成要素的第五点就突出强调，建设 e-Science，需要建设者“参与到国际网格项目和活动中”。中国与英国在 e-Science 领域的合作尤其明

显。目前，中国与英国在生命科学研究领域上的 e-Science 合作进展顺利。截止于 2006 年 12 月，中国生命科学研究院已经建立起 11 个 e-Science 网格结点，而英国则建立了 10 个 e-Science 网格节点。通过这些网格节点，两国科学家便可以实现生命科学研究数据的共享。

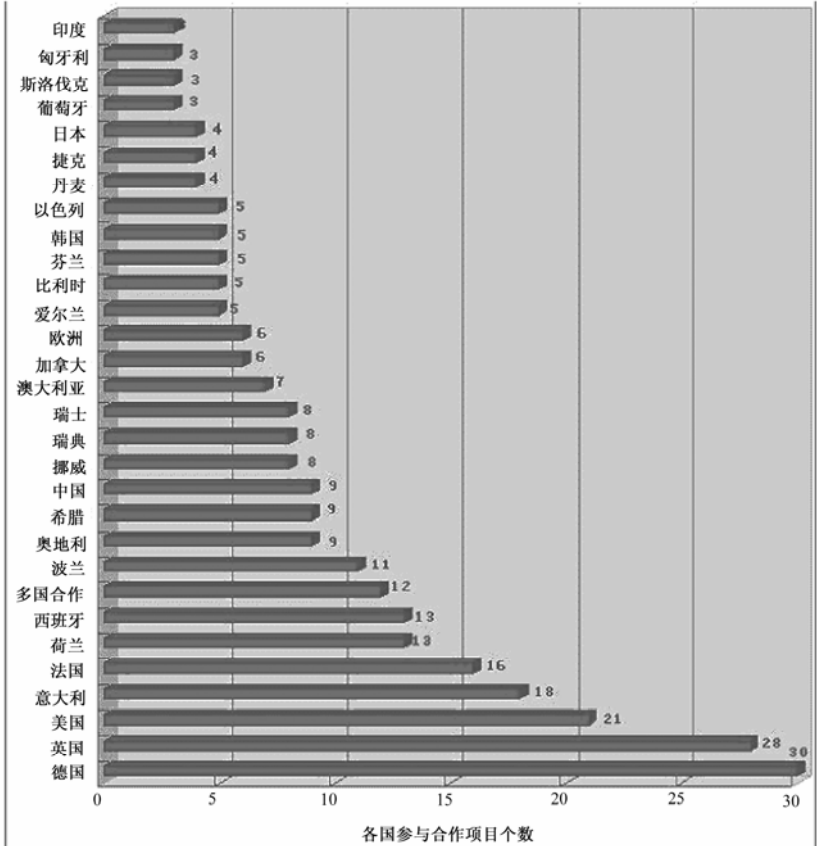


图 10-7 调查项目典型国家参与国际合作项目统计分析

美国共在 22 个 e-Science 项目中与 37 个国家进行合作，其中有 5 个项目美国是作为项目的发起者和领导者。与英国合作项目有 15 个，与荷兰和加拿大合作的项目有 5 个，与澳大利亚、韩国、日本和中国合作的项目均是 4 个。中国（包括台湾地区）与美国合作的 4 个项目，分别是 STAR、PRAGMA、ACCESS，Grid 和 BaBar，其中的 STAR 和 BaBar 均是高能物理方面的 e-Science 应用项目。

同时，美国与英国在相关领域的合作最为频繁，在美国所有的国际合作项目中，有 68.2% 的项目都有英国的参与。

欧盟启动的 e-Science 建设项目一般都是由 25 个成员国共同参与。在欧盟所参与的 40 个项目中，一个项目至少有 2 个国家参与，而最多的一个项目则有 27

个国家共同参与。在所有项目中，参与国家数目在 2~5 个之间的项目数最多，总共有 22 个，占全部项目的 54%；其次是参与国家数目在 6~10 个之间的项目，共有 15 个，占全部项目的 38%；而参与国家数目在 11~27 个之间的只有 3 个项目，占全部项目的 8%。此外，在这 40 个项目中的 7 个项目有欧盟以外的其他国家参与。

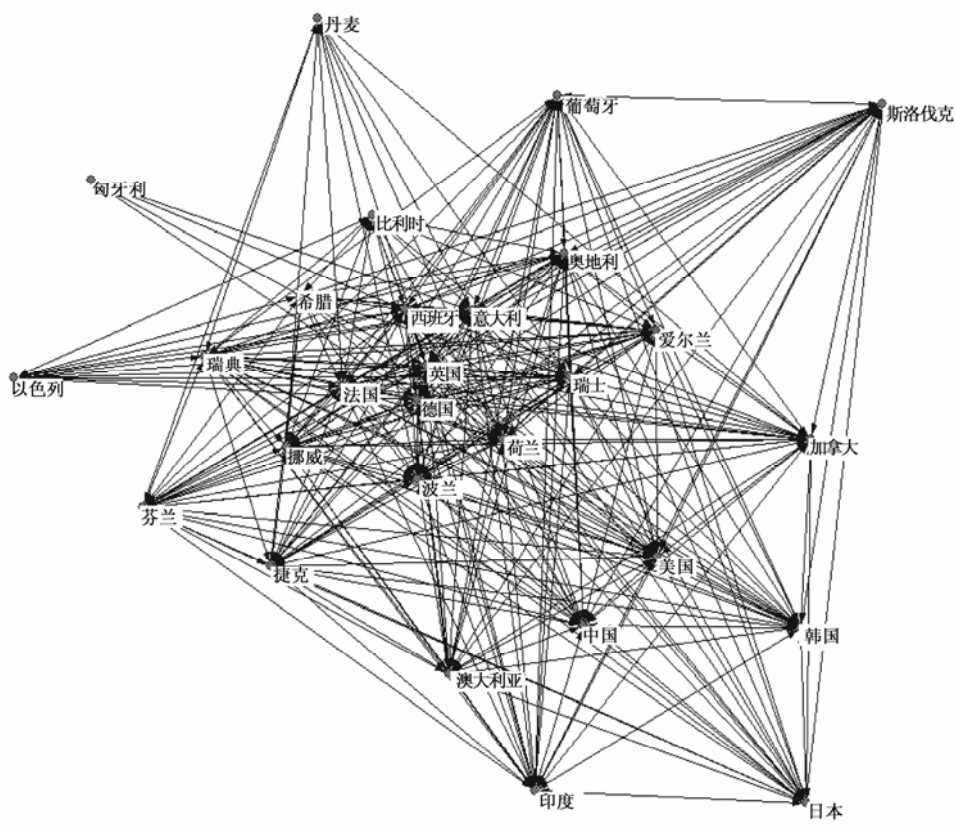


图 10-8 典型国家参与国际合作项目图

10.4.2 高校、科研机构通力合作是e-Science建设的又一特征

1. e-Science建设主要由高校与科研机构共同参与

为统计分析方便，可将调研项目中的所有承担机构划分为 6 类型：政府部门、科研机构、高校¹、公司企业、协会组织、其他机构²。全部调研项目和机构数量按机构类型的分布情况如图 10-9 所示。图中“项目数”表示由某一类型机构参与建设

¹ 高校作为一个承担机构整体，包括高校院系、高校研究中心和高校实验室等。
² 其他机构主要包括图书馆、医疗机构、教育机构等。

的e-Science项目总数，“机构数”表示某一类型机构所包含的e-Science项目承担机构数量。从图中可以看出，高校是调研项目中最大的承担机构，总共有 336 所高校承担了 371 个e-Science项目，占承担机构总数的 45%，占全部调研项目的 57%；科研机构和企业则分别排在第 2 和第 3；这三类机构总数则占到了承担机构总数的 88%，可见，仅从数目上讲，这三类机构是e-Science建设过程中的主要参与者。此外，图 10-10 统计了每种机构平均承担的项目数。从图 10-10 中可以看出，平均每所高校承担的项目最多，接近 3 个；科研机构、政府部门和公司企业平均承担的项目数也较多。

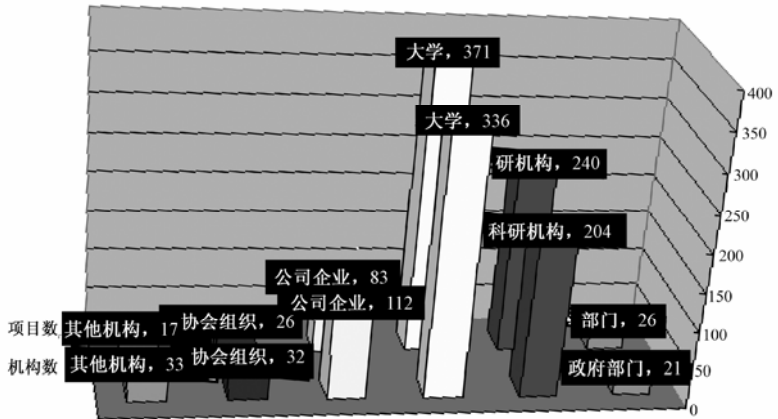


图 10-9 e-Sicence 项目和承担机构按机构类型的分布情况

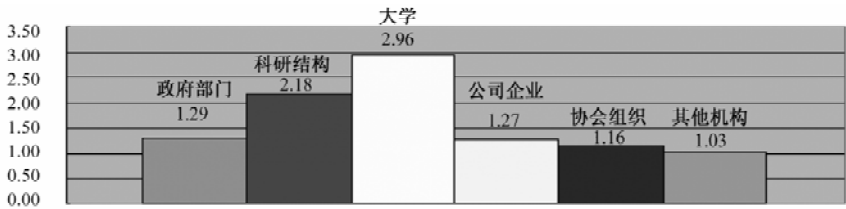


图 10-10 承担机构按类型平均承担项目数

图 10-10 是整个世界范围内项目承担机构情况分析，而各个地区在项目承担机构类型分布上又显示出了自身的特点。在调研的美国地区的 94 个项目中，共包括 160 个项目承担机构，其调研项目和机构数量按机构类型的分布情况如图 10-11。而从图 10-11 中可以看出，承担机构数排在前三位的依次是高校、科研机构和政府部门，而美国公司企业则在项目承担中相当较少。可见，美国在 e-Science 建设过程中公共研究机构扮演的角色比较重要，而公司企业参与则不够积极，这主要是因为 e-Science 虽然有着广阔的发展前景，但毕竟这是一项耗资巨大、牵扯多方因素的工程，而且不会在短期内收到更多利益上的回报，对于这样的投资大、风险大的项目，而那些

感兴趣的公司企业或者个人，也许也会积极参与，但在美国 e-Science 建设的当前阶段还不是主要参与者。此外，虽然高校是美国 e-Science 项目中最大的承担机构，但是平均每个高校承担的项目数却不高（图 10-12），造成这一现象的主要原因是，在高校这一类型中包括高校院系、高校研究中心和高校实验室 3 个部分，其中只有高校研究中心和高校实验室，以及少量的高校院系具有较强的研发能力，其他在“高校”这一类型的承担机构中占较大比重的是研发能力较弱的高校院系，这些高校院系可能只是偶尔在某些特定领域的项目研究中作为项目应用的参与者出现。

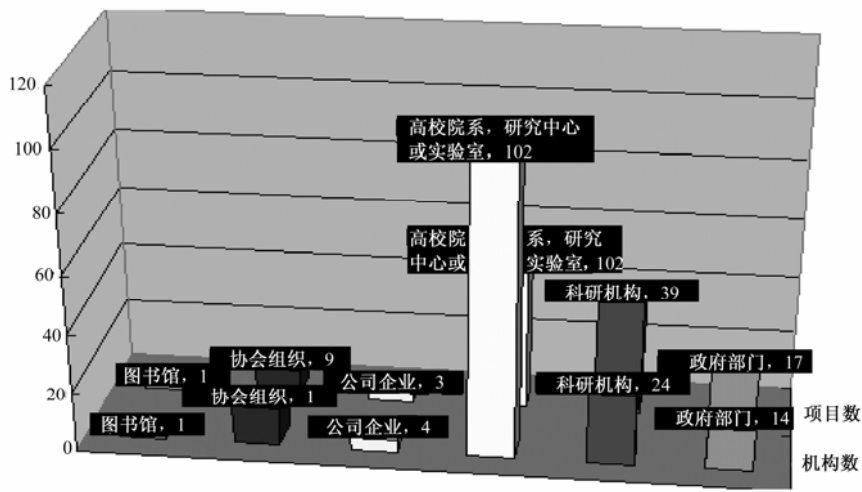


图 10-11 美国承担机构数量及所承担项目数量按类型分布

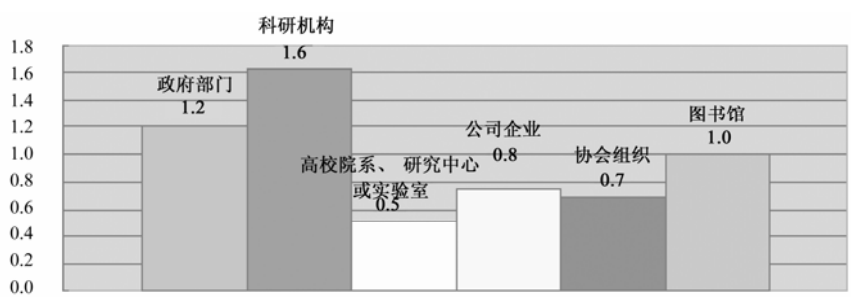


图 10-12 美国不同类型承担机构平均承担项目数量

2. 典型国家参与机构比较分析

根据项目承担机构的数量，本书编者选择了英国、美国、加拿大、德国、法国以及中国 6 个典型国家，对它们的承担机构按照上述 6 种类型划分的承担机构数量、承担项目数量以及平均承担项目数量进行比较分析，以期揭示各种类型的承担机构在世界各国 e-Science 建设中扮演的角色。

图 10-13 为上述 6 个国家承担机构数量按承担机构类型的比较情况。从图中可以看出，除法国之外，其他 5 个国家中的高校在 6 种项目承担机构类型中都是数量最多的机构类型。除了这个共性外，6 个国家在承担机构类型上都有自己的一些特点。在政府机构参与科研方面（不包括资助方面），美国比较突出；而英国和法国显然有更多的科研机构参与到了 e-Science 的建设当中；英国和加拿大在研发的过程中，较多利用了来自公司企业的力量；在协会和组织类型的承担机构方面，美国和英国在数量方面有更多的参与者；而加拿大的其他类型的机构显然比其他国家更多的加入到这一领域的研究当中，不过对于加拿大的“其他机构”较多这一事实，这里还是要作以初步解释。因为在对加拿大调研的过程当中，收录了大量 e-learning 方面的研究项目，而这方面的承担机构涉及较多的教育机构，由于归类时这部分机构被划分在“其他机构”类型当中，所以可能导致加拿大在这方面的机构数量较高。

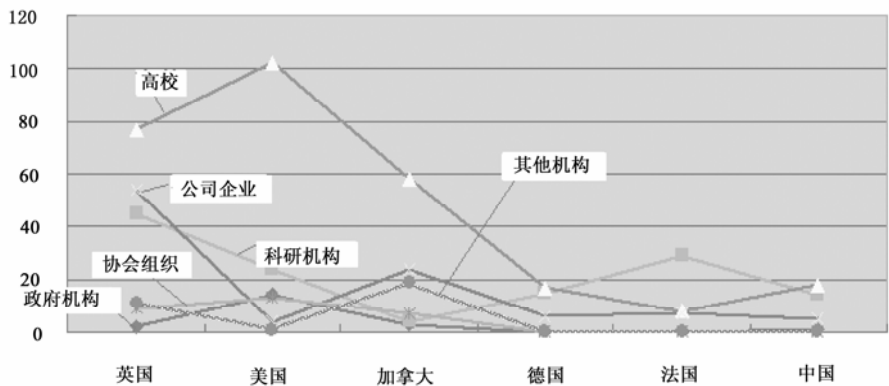


图 10-13 六国承担机构数量按类型比较

图 10-14 为 6 国不同类型机构参与项目建设的情况比较。从图 10-14 中不难看出，英国不论是高校、科研机构还是公司企业都参与到了大量的 e-Science 项目当中，而美国次之。当然，这与各地区的调研项目在数量上的差异是分不开的，因此通过使用各国机构参与的项目总数去除以各类型的机构分别承担的项目数，则会得到图 10-15 所示的六国各类型机构在 e-Science 项目中的参与程度数据。从图中可以看出，e-Science 项目中，在政府机构的参与程度方面，美国最高，其次是加拿大、中国；在科研机构的参与程度方面，中国最高，其次是美国、英国；在高校的参与程度方面，英国最高，其次是中国、加拿大、美国，这四个国家高校的参与度都超过 50%，即这些国家超过一半的项目都有高校参与；在公司企业的参与程度方面，加拿大最高，其次是英国、德国、中国；在协会组织的参与程度方面，加拿大最高，其次是美国、英国；在其他机构的参与程度方面，则以加拿大最高。

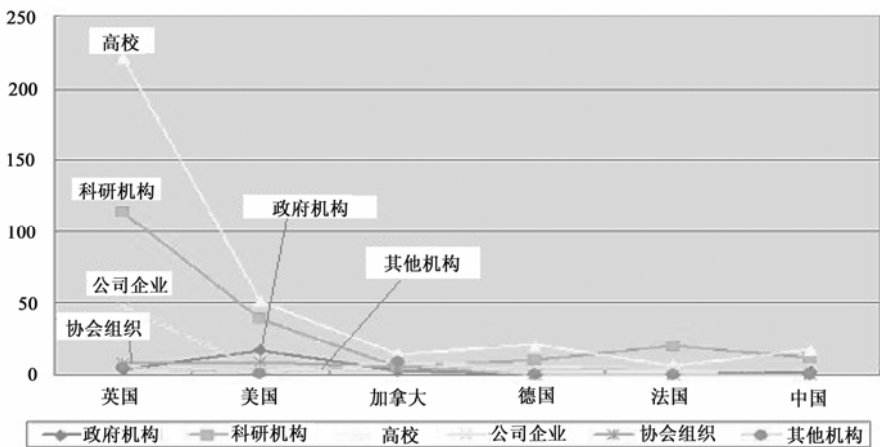


图 10-14 六国 e-Science 项目数量按承担机构类型比较

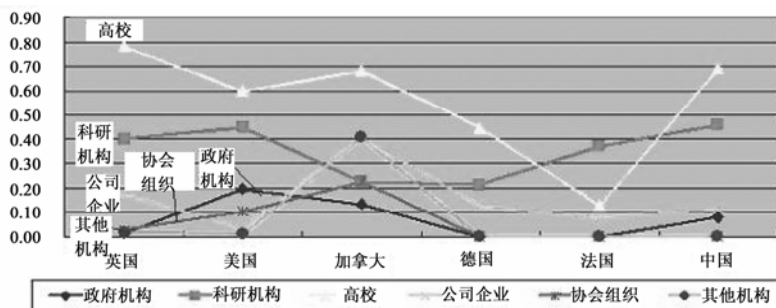


图 10-15 六国各类型机构在 e-Science 项目中的参与程度比较

此外，对于每种类型机构中平均每个机构参与项目数的比较情况如图 10-16 所示。从图 10-16 中我们可以看到，美国的科研机构和英国的高校平均承担的项目数均为 7 个左右；英国的科研机构、政府机构，美国的高校和中国政府机构平均承担 2 至 3 个项目¹；而为数众多的其他类型的机构大都是只承担过一个项目。

3. 各机构承担项目数量分析

对全部 679 个项目承担机构中每个机构承担的项目数量进行了统计，本书列出了承担项目超过 10 个的机构（如表 10-2 所示）。尽管我们不能仅从承担项目数量上评价一个机构在 e-Science 领域的重要程度，因为毕竟项目的规模和项目成果的价值才是主要的评判标准，但承担项目数量的多少至少可以说明该机构在参与 e-Science 领域研究的积极程度。从表 10-2 我们可以看到，英国的爱丁堡大学、曼彻斯特大学、

¹ 在调研项目中实际只包括了 中国地质调查局 1 个政府机构，而该机构参与了 2 个 e-Science 项目建设，分别是“环境资源网格”和“863 空间信息网格”项目。

伦敦学院大学、牛津大学和南安普敦大学，美国阿贡国家实验室、美国劳伦斯伯克利国家实验室（Lawrence Berkeley National Laboratory）都是 e-Science 项目的主要参与机构，从表 10-2 可以看到，中国的北京航空航天大学、中国科学院和清华大学三个机构的承担项目数都超过了 10 个，但是表 10-2 中的大部分机构都分布在美国和英国地区，这一事实从一个角度反应了世界范围内英国和美国在 e-Science 研究方面的繁荣程度。

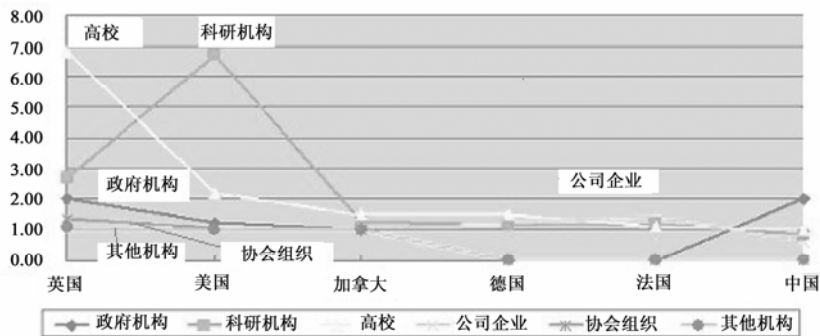


图 10-16 六国各类型机构平均承担项目数量

表 10-2 承担项目超过 10 个的机构列表

机 构 名 称	项目数	机构类型	国家
University of Edinburgh	60	高校院系	英国
University of Manchester	42	高校院系	英国
University College London	33	高校院系	英国
University of Oxford	28	高校院系	英国
University of Southampton	28	高校院系	英国
Imperial College London	25	高校院系	英国
Argonne National Laboratory	25	研究机构	美国
University of Cambridge	24	高校院系	英国
Lawrence Berkeley National Laboratory	23	研究机构	美国
Cardiff University	18	高校院系	英国
University of Birmingham	16	高校院系	英国
National e-Science Centre	16	研究机构	英国
University of Sheffield	13	高校院系	英国
Lancaster University	12	高校院系	英国
University of Glasgow	12	高校院系	英国
University of Leeds	12	高校院系	英国
University of Newcastl	12	高校院系	英国
University of Nottingham	12	高校院系	英国
North West Regional e-Science Centre	12	研究机构	英国

续表

机 构 名 称	项目数	机构类型	国家
Indiana University	12	高校院系	美国
Lawrence Livermore National Laboratory	12	研究机构	美国
Oak Ridge National Laboratory	12	研究机构	美国
San Diego Supercomputer Center	12	研究机构	美国
University of London	11	高校院系	英国
北京航空航天大学	11	高校院系	中国
King's College London	10	高校院系	英国
University of Bristol	10	高校院系	英国
Brookhaven National Laboratory	10	研究机构	美国
University of Southern California Information Sciences Institute	10	高校院系	美国
中国科学院	10	高校院系	中国
清华大学	10	高校院系	中国

此外，可通过对各地区的机构承担 e-Science 项目的数量进行了统计，以揭示各地区内 e-Science 建设的主要参与者。通过对欧盟第五框架计划的 20 个项目和第六框架计划的 14 个项目的承担机构进行统计，得到的机构数为 234 个。在所有这些机构中，按承担项目数统计排在前五位的机构分别是德国 HLRS、德国于利希研究中心（Forschungszentrum, Juelich, FZJ）、英国曼切斯特维多利亚大学（UMAN）、德国柏林 Zuse 研究所（ZIB）和德国 Fraunhofer 研究所。在前五位中有四个机构来自德国，可见德国在欧盟内 e-Science 研究中有重要作用。此外，图 10-17 给出了按参与欧盟项目数排在前 16 位的机构名称。从项目数量来说，这些机构是欧盟 e-Science 研究的主要力量。

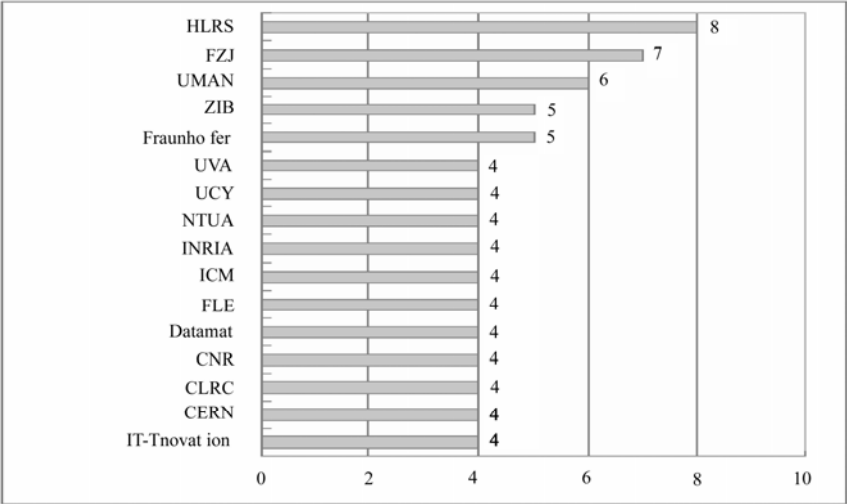


图 10-17 按参与 e-Science 项目数排在前 16 位的研究机构

统计还显示，美国地区承担项目超过 3 个（不含 3 个）的机构列表（如表 10-3 所示），通过表 10-3 我们可以看到，由 DOE 科学办公室管理和资助的 17 个国家实验室中有 12 个出现在本次调研中，在资助项目数量排名前 10 位的机构中就包括了 6 个。这一方面可能是由于本次调研选取了大量 DOE 科学办公室资助的项目，但从另一方面也在某种程度上说明了这些国家实验室在 e-Science 研究部署与应用过程中扮演了较为重要的角色。当然除了这 12 个国家实验室，还有很多科研机构 and 高校也是 e-Science 项目的重要参与者，如印第安纳大学、圣地亚哥超级计算中心、南加州大学信息科技学院等。

表 10-3 美国地区承担 3 个以上项目的机构列表

美国 e-Science 项目承担机构	承担项目数量
Argonne National Laboratory	25
Lawrence Berkeley National Laboratory	23
Indiana University	12
Lawrence Livermore National Laboratory	12
Oak Ridge National Laboratory	12
San Diego Supercomputer Center	12
Brookhaven National Laboratory	10
University of Southern California Information Sciences Institute	10
Los Alamos National Laboratory	9
National Center for Supercomputing Applications	9
University of Chicago	9
University of Wisconsin at Madison	9
Pacific Northwest National Laboratory	8
University of California at San Diego	8
Massachusetts Institute of Technology	7
Fermi National Accelerator Laboratory	6
Sandia National Laboratory	6
National Center for Atmospheric Research	5
Stanford Linear Accelerator Center	5
University of Utah	5
California Institute of Technology	4
General Atomics	4
Harvard University	4
State University of New York at Stony Brook	4
University of Arizona	4
University of California Los Angeles	4
University of Illinois at Urbana-Champaign	4

续表

美国 e-Science 项目承担机构	承担项目数量
University of Maryland	4
University of Southern California	4
University of Tennessee	4

10.4.3 项目承担机构在e-Science合作建设中角色定位各不相同

1. 各类型机构在e-Science建设技术层次中的角色

在统计分析中，我们将 679 个项目按 6 种机构类型映射到 e-Science 建设的三个层次，分别是具体应用层、应用开发环境与工具层和网格层（包括网格中间件层和网格基础设施层），从统计的角度揭示了这 6 种类型的机构在 e-Science 建设三个层次中的角色。

2. 具体应用层

在所有具体应用层的项目中，承担机构中出现最多的机构类型是高校，总共出现了 167 次，占全部具体应用层承担机构的 51%；其次是科研机构，占全部承担机构的 30%；排在第三位的是公司企业，占 12%。可见，这三类机构在 e-Science 具体应用层建设中扮演主要角色。（见图 10-18）

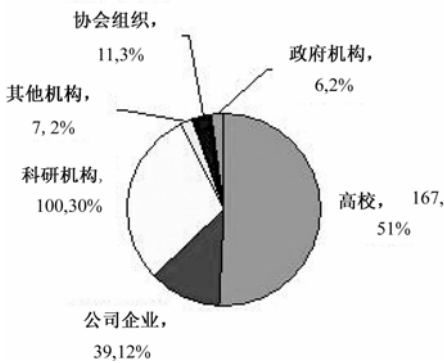


图 10-18 六种类型机构在 e-Science 具体应用层建设中的角色

3. 应用开发环境与工具层

在应用开发环境与工具层，承担机构最多的依然是高校，出现次数为 188 次，占这一层全部机构的 48%；其次是科研机构，占全部机构的 35%；排在第三位的是公司企业，占 11%。在应用开发环境与工具层，高校、科研机构和公司企业依然是主要的承担机构（见图 10-19）。

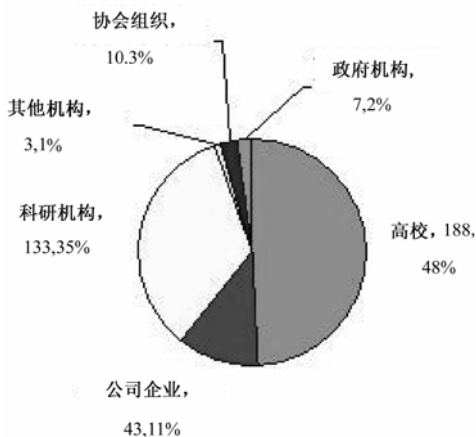


图 10-19 六种类型机构在 e-Science 应用开发环境与工具层建设中的角色

4. 网格层

在网格层中，承担机构出现次数最多的还是高校，总共出现了 199 次，占全部机构的 51%；其次是科研机构，占全部机构的 32%；第三是公司企业，占全部机构的 10%。这三类机构同样在网格层扮演主要角色（见图 10-20）。

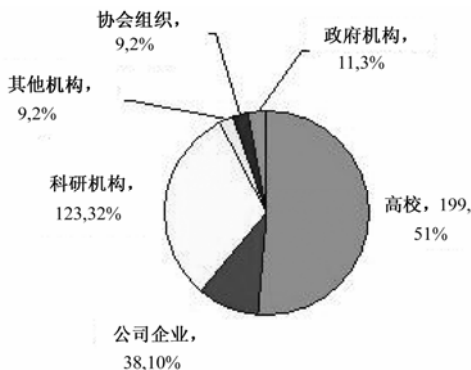


图 10-20 六种类型机构在 e-Science 网格层建设中的角色

5. 不同类型机构参与e-Science建设的层次各有侧重

在前文中，我们得出高校、科研机构和公司企业在 e-Science 建设的各层中都扮演着主要角色，本部分将进一步分析这三类主要机构在各层中的建设重点（图 10-21）。从图中可以看出，高校在 e-Science 建设层次中偏重于网格层，而科研机构和公司企业则更多地偏重于应用开发环境与工具层。

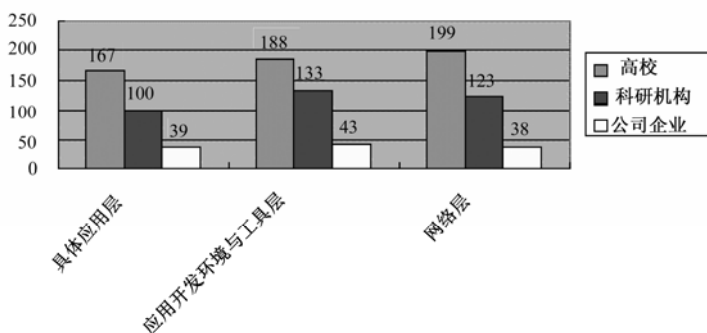


图 10-21 主要机构类型在 e-Science 建设层次中的建设重点

10.5 各国e-Science建设与学科领域、具体应用紧密结合

e-Science 建设的目标，决定了在其发展的过程中，必然是为了解决具体学科或应用领域的实际问题的。在我们的分析中，所有的 e-Science 建设项目中有 46% 的项目是面向具体学科应用进行研究和建设的。

1. 生物、医药卫生、工业技术、“大科学”是当前e-Science应用的主要领域

在 e-Science 建设项目所涉及的具体学科领域，主要包括：数学、物理、化学、天文学、地球科学、生物学、环境科学、医药卫生、计算机科学、航空航天，以及工业技术。另外还包括社会科学、历史地理、政治法律、经济、军事、语言学、教育、建筑等方面的应用。根据对调研项目的统计归类，各学科具体的项目数量的分布情况如图 10-22 所示。

从图 10-22 中我们可以看出，在生物、医药卫生、工业技术、物理、地球科学、天文学等学科汇聚了更多的研究项目，其中具体的应用如附录 H 所示。分析后便可发现，项目较为集中的应用领域，正是那些需要大量科研人员参与、进行复杂或大范围的科学试验、产生大量数据的科学研究，是那些单凭少数科学家或单个研究机构很难实现的科学研究，换句话说，就是那些需要大规模、大范围合作的科学研究，是“大科学”，而这也正是 e-Science 的目标所在，价值所在。

但各国在 e-Science 项目建设的学科应用规划方面存在一定的差异，这主要是由各国的科研政策及研究重点所决定。具体说来，如图 10-23 所示，是美国、英国（以英国 e-Science 中心的 231 个项目为例）e-Science 项目的学科分布情况。从图中我们可以看出美国更为关注地球科学、物理学方面的 e-Science 建设，而英国则更为关注生物科学、医药卫生和工程物理方面的 e-Science 建设。

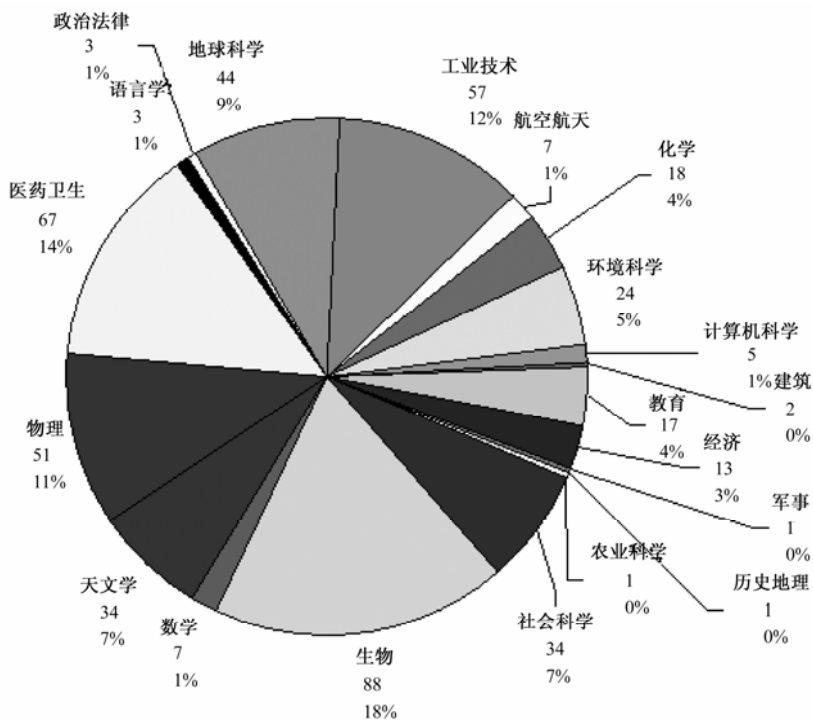


图 10-22 e-Science 项目应用学科领域分布

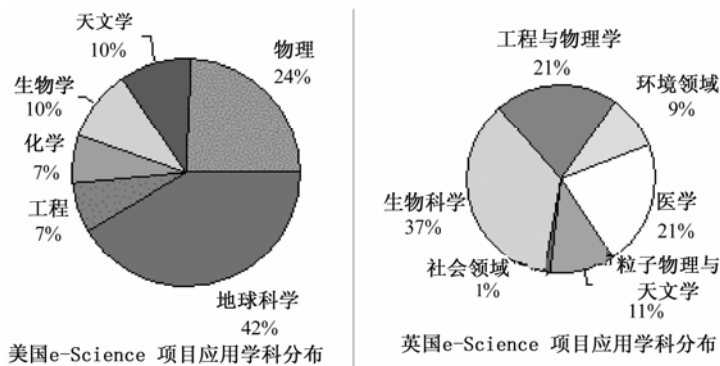


图 10-23 英美 e-Science 项目应用学科分布

2. 当前 e-Science 研究仍以基础和应用研究为主

在本次 e-Science 建设项目分析中，将调查的项目分为七种类型：基础研究、应用研究、产品研发、工具研究、试验推广、调查类和预研类，每种类型的具体含义如下。

(1) 基础研究：项目主要研究对象聚焦于 e-Science 研究中的基本概念、基础理论、算法等的研究。

(2) 应用研究：基于 e-Science 理论研究的应用性探索，重点在于探索某种方法的应用可行性。

(3) 产品研发：研发计划产品，制定产品的相关标准，并对产品的相关问题，包括对实施产品的质检、鉴定、认证等方面提出意见与建议。

(4) 工具研究：针对 e-Science 建设过程中涉及的各种工具，就工具的使用环境、性能监测、效果评估开展研究的项目。

(5) 试验推广：为了解某产品或者工具的性能或效果而进行的尝试性活动并在获得满意的效果后进行推广实施。

(6) 调查类：调查类课题主要是项目承担者根据一定的需要，为解决某一问题，确定调研方向，然后实施调查、考察，筛选调查所得的资料，通过实验与研讨，最后寻求解决问题的方法。

(7) 预研类：预研项目处于项目前期研究阶段，它是项目建设程序中十分重要的阶段，为项目资助人与直接决策者对项目的评估提供政策、技术、经济、科学的依据，帮助项目资助人与直接决策者进行项目的理性投资决策。

通过统计分析，每种类型的项目数量分布如图 10-24 所示。从图 10-24 中可以看出，研究类型的项目排在七种项目类型之首，应用研究类型的项目是最多，共有 284 个，占全部调研项目总量的 43.4%；其次是基础研究类型的项目，共有 208 个，占全部调研项目总量的 31.8%；再次是工具研究类型的项目 96 个，占全部调研项目的 14.7%；实际应用类型的项目包括试验推广和产品研发类型的，分别是 39 个和 36 个，分别占 6%和 5.5%；调查类 11 个，占全部调研项目总量的 1.7%；预研类的项目最少，只有 5 个，占 0.8%。

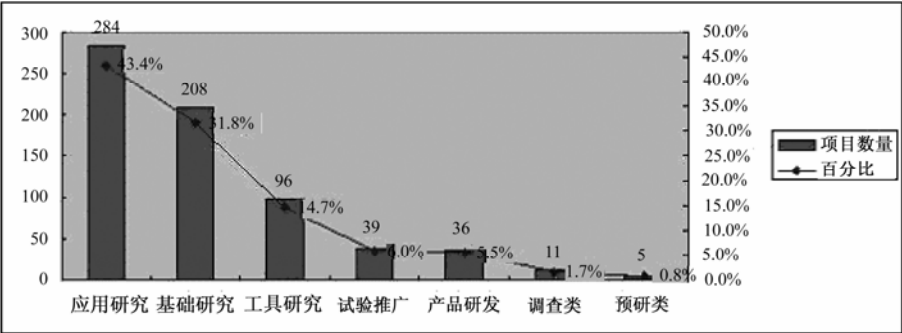


图 10-24 调研项目类型统计分析

因此，从整体情况而言，e-Science 当前的项目都比较注重于研究类，如基础研究、应用研究和工具研究，投入到实际的产品研发和实验推广都有待进一步提高。

3. 应用程序、网格系统和软件组件开发是e-Science建设项目的**主要成果形式**

根据对 e-Science 建设项目的不完全统计，总结出如图 10-25 所示的比较常见的

成果形式及对应的项目数。在统计结果中，排在前四位的分别是应用程序（68 个）、网络系统（60 个）、报告（48 个）和软件组件（48 个）。

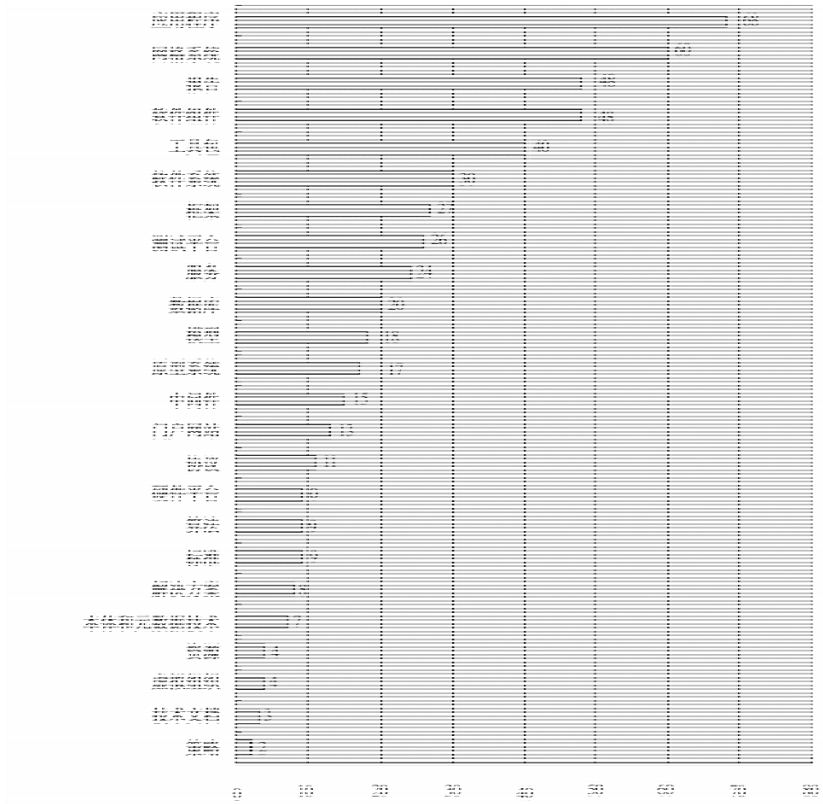


图 10-25 成果类型及对应项目数

由结果可以看出，在 e-Science 项目中，项目成果大部分是用于实际的应用，这个结果与本部分对项目类型进行的统计分析结果一致，在对项目类型统计分析中，应用研究型项目有 266 个，占有所有调研项目的 40.8%。这两个结果同时说明了当前整个 e-Science 的发展是以实际应用为主，大部分的项目都是以应用为目的。

e-Science 的发展是建立在大量的技术的基础之上的，其中网络技术是 e-Science 中具有代表性的核心技术。网络技术集成了计算机科学领域中多种技术的最新成就，代表了信息技术和信息化发展的方向，特别是在现阶段，突出体现和满足了科学研究的需要，在一定程度上，网络几乎就是 e-Science 的同义词，通过网络才可能将 e-Science 所需要的众多技术与资源，如分布式计算技术、网络安全技术、协同工作技术、资源管理技术等，集成为一体，构成并实现 e-Science。因此，在调研中，网络系统作为第二大项目成果形式，在 60 个项目上都有体现。

软件组件在 48 个项目上出现。软件组件是为自包含的、可编程的、可重用的、与语言无关的软件单元，软件组件可以很容易被用于组装应用程序中。

与前应用程序与网格系统开发的成果不一样，“研究报告”作为项目成果与软件组件同时在 48 个项目中出现。在 e-Science 项目的发展过程中，在实践应用的同时，也在不断发展理论方面的研究，当前各领域都非常重视 e-Science 的发展，对于其的理论研究（大多数以报告的形式）也在层出不穷。

总体来看，e-Science 每一个项目的成果都不是单一的。以欧盟第五框架计划的网格项目成果为例，主要体现在七个方面：① 网格标准化；② 中间件开发；③ 商业化应用；④ 科学应用；⑤ 互操作；⑥ 网格开发工具；⑦ Web 服务。GRIDSTART 委员会在总结第五框架计划下网格项目成就时具体给出了 16 个项目在这七项技术成就上的统计（表 10-4）。

表 10-4 欧盟第五框架计划网格项目的技术成就

技术成就/ 项目名称	网格 标准化	中间 件开发	商业 化应用	科学应用	互操作	网格 开发工具	Web 服务
BioGrid			•			•	•
CrossGrid		•		•	•	•	
Damien		•	•			•	
DataGrid	•	•		•		•	
EGSO				•	•	•	•
EuroGrid	•	•	•	•			
FlowGrid		•	•			•	
GEMSS		•	•				•
GRACE	•	•	•	•			
GRASP		•	•			•	•
GRIA		•	•			•	•
GridLab		•		•		•	
GRIP	•	•			•		
MammoGrid			•	•			•
OpenMolGrid			•	•		•	
WebSI		•	•				•

10.6 各国e-Science规划实施各有特色

10.6.1 e-Science规划布局模式不同

1. 英国强调顶层设计、分布实施

英国的 e-Science 建设在顶层有一个总体规划，它的设计者是英国贸工部隶属的科技办公室。而在实施过程中，科技办公室负责制定与 e-Science 密切相关的各项宏观政策，并负责各研究理事会在 e-Science 方面的拨款，而具体的实施则是在英国

e-Science 计划指导委员会的指导下，由英国研究理事总会和各研究理事会分布实施。目前，英国已经建成了一个庞大的 e-Science 研究与建设体系。这个研究与建设体系包括：一个国家 e-Science 研究中心、十个地区 e-Science 研究中心、七个优秀的 e-Science 研究中心以及七个面向专门领域的 e-Science 中心。

2. 美国以项目为核心分散实施

美国在 e-Science 建设过程中没有一个统一的规划，而是以项目驱动的方式来推进 e-Science 建设，这也是美国科研体制的一个重要特点，也是美国在 e-Science 建设走网格技术路线的必然选择。这种纯粹以项目的方式组织 e-Science 建设，给 e-Science 研究的宏观管理和整体评估造成一定的困难，但是由于各科研机构是在识别自身对 e-Science 需求的基础上启动相关的研究项目，因此，美国的 e-Science 研究成果能更好地选取和开发新技术，满足科学家数字科研的需求。

3. 欧盟采取局部试点、平移扩展方式组织泛欧洲e-Science建设

欧盟在 e-Science 建设中的一个特色是局部试点、平移扩展。欧盟 e-Science 建设首先在欧盟主要成员国内进行试点，在取得一定成就后向其他欧盟成员国进行推广，最终目标是在泛欧洲（Pan-Europe）范围内实现 e-Science 建设的整合。欧盟 e-Science 建设的这种模式以 EGEE（Enabling Grids for E-Science in Europe）和 SEE-GRID 两个项目为典型代表。EGEE 项目启动于 2004 年 4 月，该项目是 DataGrid 项目的扩展和延续。2001 年启动 DataGrid 项目时只有英国、法国、德国等 10 个国家参与，该项目于 2004 年 3 月圆满完成。欧盟为了推广 DataGrid 项目的建设经验，将 e-Science 建设成果扩展到其他欧洲国家，在 4 月即启动了 EGEE 项目，该项目总共有 27 个国家参与，EGEE 项目的完成必将会使整个欧洲的 e-Science 建设上一个台阶。SEE-GRID 项目是 e-Science 建设平移扩展的典型，该项目是东南欧地区的网格基础设施项目，旨在将欧盟 e-Science 建设的成果和经验平移到东南欧地区的 10 个国家，从而使 e-Science 建设在泛欧洲范围内广泛展开。这种平移扩展机制主要包括开展 e-Science 建设经验推广会、发放培训材料（包括各种项目手册、试点计划和示范项目、测试平台的资料）等，并成立运行和支持中心，开展可行性研究，开发 e-Science 建设路线图，并最终希望通过泛欧洲的网格基础设施将东南欧地区整合进欧盟研究区（European Research Area）。

10.6.2 e-Science建设具有明显的技术层次性

欧盟 e-Science 建设已经进行了五年之多，欧盟 e-Science 建设从时间上看是分两个阶段来进行的，即第五框架计划（1998-2002 年）是第一个阶段，第六框架计划（2002-2006 年）是第二个阶段。按照 e-Science 建设的五层框架体系来审视欧盟

e-Science 的建设过程，我们可以发现欧盟在不同的时间阶段突出建设的层次是不同的。欧盟 GRIDSTART 项目将第五框架计划下的项目聚成两类，分别为第一波项目和第二波项目，研究发现第五框架计划第一波资助的 e-Science 项目集中在五层框架的较低层次，其中大部分项目都集中在网格基础设施和网格中间件两个层次，而且这一阶段的资助特点是资助项目数量少但资助金额大；而第五框架计划第二波资助的网格项目则集中在五层框架的较高层次，主要是开发特定领域内的 e-Science 应用，这一阶段欧盟资助的特点是资助项目数量多但资助金额小（图 10-26）。第六框架计划资助的 e-Science 项目则继续集中在五层框架的较高层次，可见，欧盟的 e-Science 建设已经从最初的基础设施建设过渡到面向应用领域的建设，这种过渡也说明欧盟早期在 e-Science 基础设施方面的建设已经为 e-Science 应用研究提供了一个很好的基础。

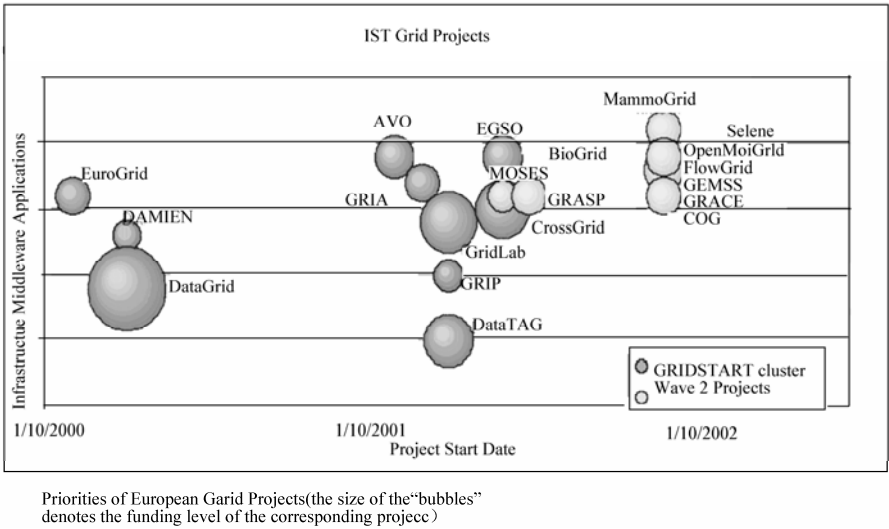


图 10-26 欧盟网格项目的特点

3. 建设中不同参与者承担不同层次的角色

在对 679 个项目调研中，根据五层结构中的具体应用层、应用开发环境与工具层和网格层（包括网格中间件层和网格基础设施层）按照国别进行机构分类（由于 e-Science 与各种具体应用密切结合，而资源层由具体应用决定。因此，考虑到其复杂性，本课题并未对资源层进行分类），各国参与机构在各层中发挥的作用各有特色，本课题主要选取了中国、英国、美国、加拿大、法国和德国六个国家进行分析。在调研项目中，将承担机构划分为 6 种类型：政府机构、科研机构、高校（包括高校院系、高校研究中心、高校实验室）、公司企业、协会组织、其他机构（如图书馆、医疗机构、教育机构等）。每个项目的承担机构在各层中出现频次可以从一个角度反映出一个国家中各机构在五层机构中的具体作用。

5. 中国科研机构在网格层和应用开发环境与工具层发挥重要作用

中国各机构在五层机构中的角色发挥的作用如图 10-27 所示。主要参与机构是科研机构、高校、公司企业和政府部门。高校在各层都发挥着积极的主导作用，其中在应用开发环境与工具层以及具体应用层都作出了巨大的贡献，如高校作为参与机构在应用开发环境与工具层出现了 86 次，具体应用层出现了 89 次；科研机构在五层机构中也比较活跃，涉猎各层，但是作用要稍逊于高校；此外，图 10-27 也表明，政府部门也参与到应用开发环境与工具层和具体应用层的建设中。

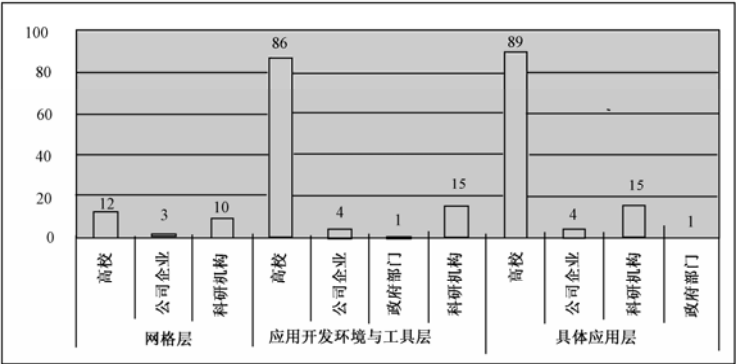


图 10-27 中国各机构在五层结构中的角色发挥

由图 10-27 可以得出，在国内 e-Science 的建设过程中，社会广泛参与，高校、科研机构和公司企业积极参与五层结构中的建设，其中主要是以高校和科研机构为主体，而公司企业和政府部门广泛参与。

3. 英国的高校在e-Science建设中居于绝对领导地位

e-Science 在英国的发展可谓是全民动员（如图 10-28），六种类型的机构在各层都发挥着他们自身积极的作用。其中高校在这六种机构类型中以绝对的优势，发挥着积极的主导作用；其次是科研机构。政府部门在各层也发挥着自身的作用。

4. 高校和科研机构是美国e-Science建设的主体

美国各机构在各层也比较活跃（详见图 10-28），基本各层都有各机构的身影，尤其是在应用开发环境与工具层，六种类型的机构均广泛参与。高校、科研机构、协会组织和政府部门参与了所有层的建设。在各层的参与活动中，科研机构和高校的作用几乎并驾齐驱，不相上下，科研机构在具体应用层和网格层的作用最大，高校在应用开发环境与工具层的作用最大。

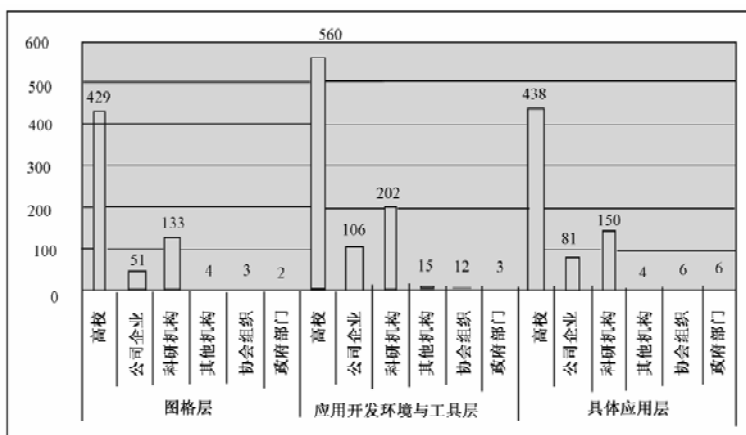


图 10-28 英国各机构在五层结构中的角色发挥

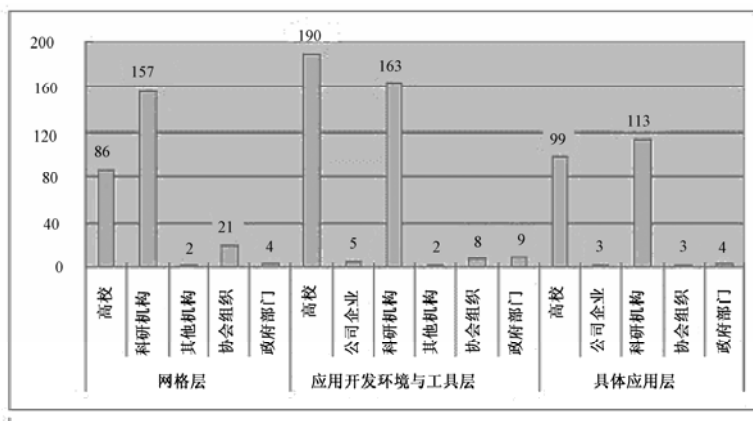


图 10-29 美国各机构在五层结构中的角色发挥

4. 加拿大的高校在e-Science建设中主导网格技术研究

加拿大各机构广泛参与 e-Science 建设 (如图 10-30 所示), 基本各层都有各机构的身影, 其中网格层涵盖了六种类型的机构, 高校承担网络建设的主要任务。高校、公司企业、政府部门参与了所有层的建设。各类型的机构在各层的作用都比较均衡, 势均力敌。其中, 高校在网格层的作用比较突出。

5. 法国临床医疗与教育机构主导应用开发环境与工具层和具体应用层的建设

法国当前 e-Science 建设涉及的应用领域主要集中在临床医疗与教育方面。因此, 在建设的过程中, 医疗机构与教育机构表现积极, 而且较多地参与到各类项目的网格层建设中。为方便统计, 本课题将医疗机构和教育机构等一并归入到了“其他机构”这一类别中, 从图 10-31 可得出, 与具体应用层、应用开发环境与工具层参与的

机构数相比，具体应用层中的其他机构数量最多，一共有 62 个医疗机构、教育机构或者是与医疗教育有关的机构参与其中的建设。

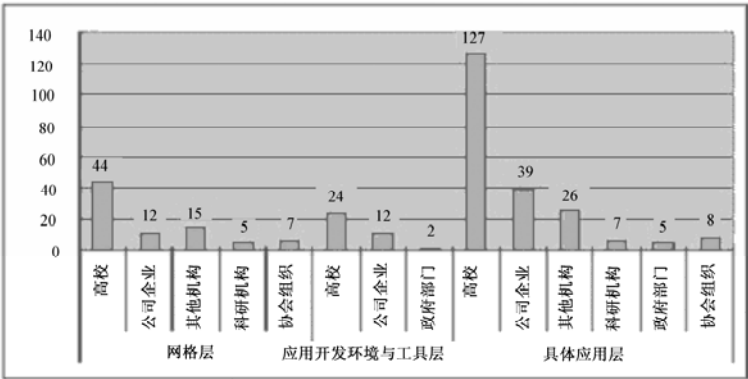


图 10-30 加拿大各机构在五层结构中的角色发挥

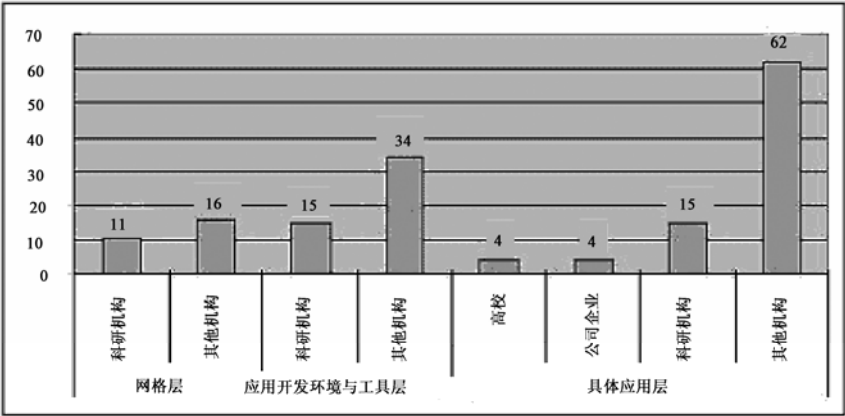


图 10-31 法国各机构在五层结构中的角色发挥

6. 德国社会各界广泛参与e-Science建设

与法国情况类似（如图 10-32），德国在 e-Science 的项目开展中，各类型机构的作用也都不是特别突出，e-Science 建设的重点也相对集中于医疗和教育领域，科研机构、高校和公司企业都积极参与。其中，科研机构和高校的作用是六种类型机构中相对突出的。

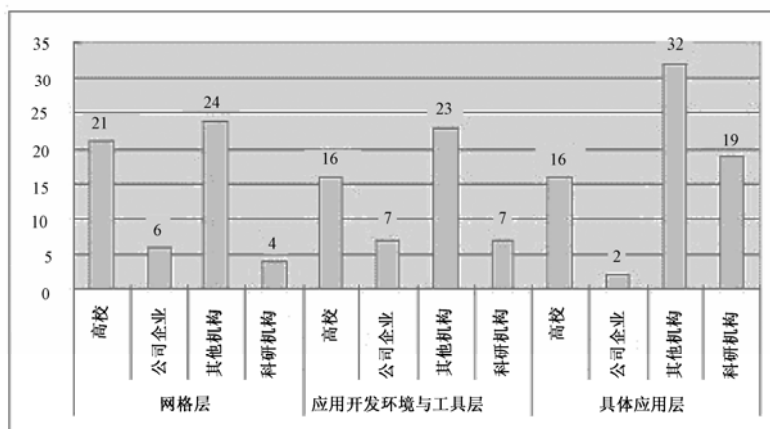


图 10-32 德国各机构在五层结构中的角色发挥

10.7 政府在e-Science建设的宏观规划中发挥主导作用

就世界范围内各国家、地区 e-Science 项目的开展状况而言，当前，各国政府在 e-Science 宏观发展规划中发挥着主导作用。

1. 英国政府主导英国e-Science规划的制定

由于各学科的 e-Science 研究和建设除了有自身的特性外，还有许多共性问题，即对所需要的基础设施有相同的要求。为了避免各学科重复研究，英国政府制定了 e-Science 核心计划，以建立适于各学科 e-Science 建设的通用技术。依照此核心计划，英国将在 3 年时间内投资 1200 万英镑建成下一代 IT 基础设施，支持 e-Science 及商业应用，以确保英国在全球网格技术研究中与美国和欧盟一同处于世界领先地位。

为指导全国 e-Science 建设工作，英国政府推动实施了 e-Science 核心研究计划。政府在爱丁堡大学和格拉斯哥大学建立了英国国家 e-Science 中心，目的在于协调和支持十个地区 e-Science 中心，促进 e-Science 技术产业化和商业化；同时负责管理建立于爱丁堡的国家 e-Science 研究院（e-Science Institute）；此外，还负责组织英国的国际 e-Science 交流活动，其目标是力争使英国成为一个全球 e-Science 研究中心。

为了保证核心计划的顺利实现，国家 e-Science 中心将核心计划分为两个阶段进行。第一阶段是 2001 年到 2004 年（已经完成），建设重点是，提出网格的研究和开发计划，支持 e-Science 测试平台。第二阶段从 2003 年开始到 2006 年结束制订长期研究目标，突出网格中间件、可持续运行的网格服务环境、网格支持环境中心、数据维护中心等方面的研究和建设。之后，英国将进入 e-Science 的普及推广与巩固应用阶段。

通过实施 e-Science 核心计划,英国政府期望,让来自世界各国的科学家能够以一种前所未有的高效率工作方式,协同式地开展科学研究,并在此过程中实现数据、设备、仪器、知识等多种资源的共享。总之,在英国 e-Science 建设的过程中,无论是核心计划的制定、实施,还是 e-Science 十大研究中心、七大优秀研究中心等机构的设立,都离不开英国政府的统筹规划。

2. 欧盟委员会主导欧盟e-Science规划的制定

欧盟委员会下的信息社会和媒体总理事会在欧盟 e-Science 建设的规划中起主导作用。该总理事会的使命是支持欧洲范围内信息通信技术的开发和利用。该总理事会下共有 10 个分理事会,其中 e-Science 相关的研究活动主要由分管“新出现的技术和基础设施”这个主题的 F 分理事会负责。而欧盟 e-Science 建设的具体规划则体现在欧盟以四年为周期的框架计划中,通过该计划,可以了解到,欧盟对 e-Science 的目标定位是,旨在使整个欧洲地区的研究者更好地分享知识和各种资源,使网格真正支撑欧洲 ERA,网格将会整合整个欧洲的世界一流的科技资本以增强欧洲的竞争力和生活质量。在第五框架计划下的项目较多地集中在 e-Science 五层结构的中下层,主要是网格基础设施层和网格中间件层。而在第六框架计划下的项目则较多地集中在五层结构的上层,更多地面向应用领域的建设。可见,欧盟委员会在 e-Science 建设的不同阶段都起着主导作用。

3. 美国政府主导美国e-Science规划的制定

尽管美国的 e-Science 研究并没有一个顶层的管理运营机构进行统一协调,但是美国自然科学基金会(NSF)和国家能源部(DOE)两大政府机构通过其大规模并且资金数额巨大的资助计划,在某种程度上主导了美国全国范围内整体的 e-Science 规划。其中,较为典型的 NSF Middleware Initiative (NMI)、Information Technology Research for National Priorities (ITR)、DOE SciDAC 和 DOE National Collaboratories 等资助计划,都对 e-Science 的结构框架、技术平台、应用领域等角度进行了具体的界定,并通过资金分配的方式突出了重点的研究和建设方向,并保证了美国 e-Science 建设在全面性和在世界范围内的先进性。

美国科研资助系统严格的审查制度和交流机制,也在某种意义上保证了 e-Science 研究的质量和研究成果的传播以及加快试验系统产业化、实用化的步伐。

因此,从总体上看,美国的政府机构不仅在 e-Science 的资助金额占据着绝对主要的地位,而且在对 e-Science 的发展与规划方面也是核心的引导和管理力量。

4. 亚洲各国主导各自国内e-Science规划的制定

中国政府高瞻远瞩,早在 2002 年,跟随英国之后,中国政府即启动了 CNGrid 建设项目。在此基础上,相继投入 NSFC Grid、Shanghai Grid、863 空间信息网格以

及中国教育科研网格等大型网格项目的建设，并逐渐形成以国家科技基础条件平台建设为顶层规划，以五大网格为先导的中国 e-Science 建设局面。与中国紧邻的日本也于 2001 年启动了 ITBL (IT Based Laboratory) 项目，在政府的统一规划下，它的目标定位为，通过信息技术来构建一个虚拟的共同研究环境。为实现预期效果，日本政府提供大量的资金援助，并规划在五年内为该项目投入 1 亿 6 千万欧元。韩国则从 2001 年开始，由政府出面规划 e-Science 建设，并在国家层面上积极与美国、英国、日本等发达国家以及中国等发展中国家达成合作，共同推进 e-Science 建设。

10.8 政府在e-Science建设的管理中发挥主导作用

通过世界主要国家和地区 e-Science 建设的管理运行机制综合分析后，发现在 e-Science 建设过程中存在两种不同的管理运行模式：集中式管理运行模式与分散式管理运行模式。前者以英国和欧盟为代表；后者则以美国为典型。

10.8.1 以英国和欧盟为代表的集中式管理运行模式

e-Science 建设的集中式管理运行模式是指一个国家或地区在 e-Science 建设中有统一的战略规划，并设定了统一的战略目标。集中式管理运行模式往往通过成立统一的 e-Science 管理机构来协调、监督、管理一国或一个地区内各研究机构的 e-Science 研究与建设活动，负责制定、实施 e-Science 研究项目的统一资助政策，并集中对 e-Science 项目的研究成果进行评估以及成果的商业化推广。

集中式管理运行模式以英国和欧盟为代表。英国在 2001 年启动 e-Science 计划时，即是由英国贸工部下的科技办公室负责统一管理，制定与 e-Science 密切相关的各项宏观政策，并负责英国各研究理事会在 e-Science 方面的统一资助，并由研究理事会总会负责协调七大研究理事会在 e-Science 研究过程中的各项活动，而具体事务的管理则是由 EPSRC 代表其他所有六个研究理事会行使。此外，英国成立了 e-Science 计划指导委员会，负责对英国 e-Science 核心计划进行直接监督和管理。欧盟的 e-Science 建设也是一种集中式管理运行模式，欧盟 e-Science 研究活动由欧盟委员会下信息社会和媒体总理事会负责统一资助管理，具体的项目管理运行则由总理事会下的 F 理事会负责。

1. 英国e-Science管理运行机制

自英国科学家于 2000 年提出 e-Science 概念以来，短短几年内，e-Science 便在世界范围内取得迅速发展。为了保证本国或者本地区 e-Science 建设的协调发展，以英国、欧盟、美国等为代表的一些国家与地区，纷纷采取相应措施，构建了与 e-Science 研究与建设特点相结合的管理运行机制。

(1) 英国 e-Science 管理运行体系

事实上，英国 e-Science 建设计划是由英国贸工部隶属的科技办公室代表英国政府推出的一项基础性科学研究计划。在实施过程中，科技办公室负责制定与 e-Science 密切相关的各项宏观政策，并负责各研究理事会在 e-Science 方面的拨款。尽管研究理事会总会负责协调七大研究理事会在科学研究过程上的各项活动，但在此次 e-Science 计划中，实际事务的管理大权却由工程与物理科学研究理事会（EPSRC）代表其他所有六个研究理事会行使。

为了确保 e-Science 建设的顺利开展，在工程与物理科学研究理事会的倡导下，英国成立了 e-Science 计划指导委员会，并由指导委员会负责对英国 e-Science 核心计划直接进行监督、管理。图 10-33 揭示了英国 e-Science 建设的组织管理关系。

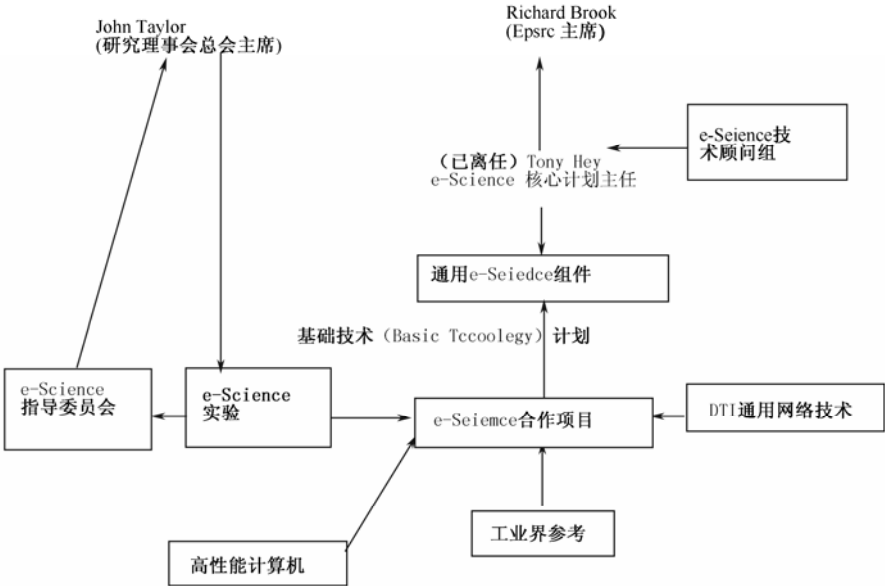


图 10-33 英国 e-Science 建设的组织管理关系图

从整体上分析，英国 e-Science 计划分为英国研究理事会总会的 e-Science 计划与各个研究理事会的 e-Science 计划两大部分。其中，研究理事会总会的 e-Science 计划又被称为英国 e-Science 核心计划，由工程与物理科学研究理事会（EPSRC）代表其他六个研究理事会负责实施，其主要任务是建设发展英国 e-Science 需要的通用组件与基础设施，在英国贸工部的支持下，e-Science 核心计划投资建立了十个 e-Science 研究中心。

在建设与研究的过程中，英国试图在全国范围内建立起发展 e-Science 需要的信息基础设施，当前，这种信息基础设施包括了英国开放基础设施研究院（OMII）、国家数据掌握管中心、国家 e-Science 中心以及英国自然环境研究理事会（NERC）资

助的 DataGrid 项目。

① 英国开放基础设施研究院 (OMII): 英国开放基础设施研究院的发展目标是, 使自身成为一个存储中心和软件工程资源库, 以便给英国 e-Science 计划开发的开源网络中间件提供支持。

② 国家数据掌管中心: 国家数据掌管中心与英国 e-Science 核心计划和英国联合系统委员会 (JISC) 开展合作, 并就数据掌管、数据保存和开放获取等重要问题共同进行探讨与研究。

③ 英国 e-Science 中心: 在欧盟资助的研究基础设施项目 EGEE 中, 英国 e-Science 中心被推举为欧洲范围内 e-Science 各类培训活动的中心, 在传播 e-Science 专业知识方面发挥着引领作用。同时, 为了给英国所开展的 e-Science 国际跨学科专题讨论会提供资源支持, 它还创办了英国 e-Science 研究院。

④ DataGrid: DataGrid 是由英国自然环境研究理事会 (NERC) 资助的项目。该项目的建设目标是提供网格服务, 以帮助科研人员快速方便地发现、传递大气数据和海洋数据。它使用美国地球系统网格 (US Earth System Grid) 项目开发出来的技术, 并在遵循国际 OAI 元数据收割协议的基础上, 开发出了支持快速数据发现的元数据架构。

(2) 英国 e-Science 管理机构

① EPSRC 的角色

工程与物理科学研究理事会 (EPSRC) 代表代表其他六个研究理事会掌管核心计划的经费。工程与物理科学研究理事会将通过国际范围的竞聘方式任命一位 e-Science 主任 (现任职于微软的 Tony Hey 曾担任此职) 负责此项工作。e-Science 核心计划主任通过一个指导委员会向英国科学技术办公室研究理事会总司长和研究理事会会首席执行官汇报。

② e-Science 指导委员会 (e-Science Steering Committee)

为协调全国 e-Science 研究项目, 英国专门成立了一个 e-Science 项目指导委员会。该指导委员会由来自学术界、工业界以及国际专家组成, 其目的是评价所有 e-Science 项目, 以确保英国的 e-Science 研究能满足各方面的需要。该指导委员会有权直接向研究理事会董事会汇报。

e-Science 指导委员会职责: 英国 e-Science 指导委员会的职责包括, 在与 e-Science 密切相关的国际国内活动上给英国 e-Science 核心计划主任提供明确指导; 尊重 e-Science 核心计划主任的意见, 审批与 e-Science 核心行动有关的基础框架; 批准核心计划中各种项目 (尤其是研究理事会资助的项目) 之间在框架内的战略优先权以及合作机会; 确保在整个 e-Science 计划进行协调、一致的安排; 同意 e-Science 核心计划主任在一致、连贯的运行计划内以具有创新性的方式开展工作; 评估 e-Science 核心计划主任、研究理事会提交的、有关 e-Science 计划实施进展的报告。

e-Science 指导委员会成员组成: 英国贸工部、英国七大研究理事会各选出代表

一名加入指导委员会，每名代表负责对各自所属的研究理事会或者是政府部门在 e-Science 方面的活动进展进行报告，并对负责反馈指导委员会提出的建议。其中，至少要有五位委员是来自学术界和企业界，以便将 e-Science 建设过程中所取得的成果向全国乃至全世界范围内加以推广。e-Science 计划的主任理所当然是指导委员会的一名成员。具体来说，e-Science 指导委员会的成员组成包括以下几个方面：领导班子（英国 e-Science 指导委员会当前设置主席一名，秘书一名）；与 EPSRC 进行沟通协调的三位联系人（由 e-Science 核心计划主任、核心计划副主任、核心计划主任助理组成）；e-Science 技术顾问组（负责监督管理 e-Science 核心计划并为 e-Science 核心计划主任提供建议。技术顾问组的职责是负责确定哪些技术是能够适用于所有研究领域的通用技术，以便优先发展。同时，它还负责新的信息技术的研发，包括即将在互联网基础上建立的数据网格技术。由 e-Science 核心计划的正副主任、技术顾问组秘书、核心计划副经理等相关人员组成）；学术成员（由国家 e-Science 中心主任、国家 e-Science 研究院院长，目前由国家 e-Science 中心主任兼任此职）等组成。另外，e-Science 指导委员会还包括有企业界成员、国际顾问、常务委员等。

③ 完整的 e-Science 建设体系

在英国 e-Science 指导委员会的直接管理下，英国设立了一个国家 e-Science 中心、多个地区 e-Science 中心以及面向各个专门领域的 e-Science 中心。截止于 2005 年 11 月，英国已经建成了一个庞大的 e-Science 研究与建设体系。这个研究与建设体系包括：一个国家 e-Science 研究中心、十个地区 e-Science 研究中心七个优秀的 e-Science 研究中心以及七个面向专门领域的 e-Science 中心。

七个面向专门领域的 e-Science 中心是：存取网格支持中心（Access Grid Support Centre (AGSC)）、数据掌管中心（Digital Curation Centre (DCC)）、国家网格服务中心（National Grid Service (NGS)）、国家 e-社会科学中心（National Centre for e-Social Science）、国家文本挖掘中心（National Centre for Text Mining (NaCTeM)）、国家环境 e-Science 研究院（National Institute for Environmental e-Science (NIEeS)）、开放中间件基础设施研究所（Open Middleware Infrastructure Institute (OMII)）。

（3）英国国家 e-Science 中心的管理机制

① 2003 年之前英国国家 e-Science 中心的机构设置

英国国家 e-Science 中心设立主任一名，直接向 e-Science 指导委员会负责，由工程与物理科学研究理事会代表所有研究理事会在国际范围内聘任。除此之外，在 2003 年之前，国家 e-Science 中心为了确保 e-Science 计划的高效实施，还专门设置了三个工作组对 e-Science 项目的日常工作提供直接指导。这三个工作组是：

设计师工作组（Architect Task Force Team (ATF)）：由英国 e-Science 核心计划资助建立，其工作目标在于，开发一个面向英国网格中间件的基础架构模型。

网格网络组（Grid Network Team (GNT)）：由英国 e-Science 核心计划资助建立，其目标是，从计算机工程的角度给英国 e-Science 计划主任以及技术顾问工作组提供

科学合理的建议与意见。

英国数据库工作组 (UK Database Task Force (DBTF)): 英国数据库工作组又称为数据库存取与集成服务工作组, 其目标就是协调各类能够为数据库存取与集成提供支持的软件组件开发活动, 因此, 对各种能够保证当前实施自动化管理的数据库可以在网格设置内继续使用的方法从一致性和有效性方面加以鉴定便成为该工作组的工作重点。

② 2003 年之后英国国家 e-Science 中心的机构设置

2003 底, 国家 e-Science 中心对各个工作组进行改动、调整, 并设置为目前的四个工作组用于指导 e-Science 活动。

基础架构工作组 (Architecture Task Force (ATF)): 负责开发出发展英国网格中间件需要的基础架构模型。自成立以来, 它已经在 OGSA 基础架构的开发方面发挥了积极作用, 并为网格技术的长远发展提出了多项富有建设性的意见。目前, 它正致力于研究英国 e-Science 基础架构中的长期基础架构问题, 深入评估英国 e-Science 计划在诸如开放网格服务基础架构等领域的投资, 同时并研究诸如普适计算等方面的技术开发问题。

网格工程工作组 (Grid Engineering Task Force (GETF)): 为英国 e-Science 原型项目的建设、测试和演示提供指导。其成员由英国十个地区 e-Science 研究中心和网格研究中心遴选产生。目前, 它正着手对英国十个地区 e-Science 中心以及部分优秀 e-Science 研究中心所开发的、面向大范围测试平台组成资源的中间件进行评估。

安全工作组 (Security Task Force (STF)): 负责识别网格开发过程中遇到的安全问题, 并对各类安全问题的主次轻重进行排序。它已经为英国 e-Science 核心计划提交了一份安全政策并正在着手研制英国 e-Science 发展需要的技术路线图。

可用性工作组 (Usability Task Force (UTF)): 负责与使用者进行交流, 确定英国 e-Science 计划开发出来的各类组件的可使用性。

用户工作组 (UK e-Science User Group): 负责检查英国 e-Science 计划各个组件, 确定各组件在满足科研人员需求方面的发展潜力。

(4) 英国 e-Science 双重交流机制

在开展 e-Science 研究与建设的过程上, 英国 e-Science 主管部门及科研人员非常注重科研成果的交流与共享, 并确定了虚拟交流与实地交流的双重交流机制。

① 虚拟交流

英国研究人员积极参与到 GGF (Global Grid Forum, 全球网格论坛) 的各项活动中。GGF 是有意仿照网际网络工程特别工作小组 (The Internet Engineering Task Force IETF) 的模式建立的, 其中具有相似的工作组, 文件处理过程, 以及组织结构。就像 Internet 依赖于 IETF 发布的标准一样, GGF 的创立者们相信中间软件标准 (Middle Software Standard) 是网格运算技术发展的基础, 目前, 英国 e-Science 研究人员已经通过 GGF 成立了二十个工作组和二十三个专题研究小组。

② 实地交流

国家 e-Science 研究院 (e-Science Institute): 英国国家 e-Science 研究院由设立在爱丁堡的国家 e-Science 中心创办。自成立以来, 该研究院已就涉及多个学科领域的所有跨学科 e-Science 建设项目召开过多个专题研讨会, 并获得来自全世界 e-Science 研究人员的积极回应。过去三年来, 由它发起的各项活动已高达 114 起, 其中包括: 7 场项目会议、8 场研究会议、34 个专题研究会、18 次培训活动、26 次高峰会议以及 6 次国际交流会议等, 同时, 它还为十位来自其他国家的科学家提供访学机会。

e-Science 年度专题会议 AHM (All Hands Meeting): 为了分享各学科领域 e-Science 项目的经验, 建立英国 e-Science 科研团体的联系, 创造一个探讨和共享 e-Science 理念的论坛, 在英国 e-Science 核心计划领导下的指导委员会组织了面向 e-Science 的专题研讨会。该专题研讨会每年举办一次。

首次研讨会 2002 年 9 月在英国谢菲尔德举行, 主题是“认识 e-Science”。以“传播 e-Science”为大会主题的第二次研讨会于 2003 年 9 月在诺丁汉举行; 而在 2004 年召开的专题研讨会更是越来越受到英国全国范围内 e-Science 建设者与理论研究者的重视, 这次以“e-Science——开阔视野”为主题的研讨会, 共收到论文 250 篇, 会议持续期间, 有将近 100 多场专题报告和小型研讨会。

在第四次 e-Science 专题研讨会上, 不仅来自英国国内的 e-Science 研究人员获得了最为密切的同行交流, 就是来自中国的 e-Science 研究人员也以最为直接的交流方式开阔了自身的研究视野。通过每年召开一次的 e-Science 专题会议, 旨在推进各行各业通过 e-Science 促进国家科技及应用创新, 可以说, 此次英国 e-Science 年度专题会议充分实现了这一目标。

(5) 英国 e-Science 安全管理机制

针对英国 e-Science 的项目如洪水般地出现, 为了方便管理, 也为了规范英国国内的各 e-Science 项目的建设, 更为了将责任分摊落实, 于 2004 年 4 月英国国家数字科研中心出台了英国 e-Science 安全管理政策。该政策从制定背景、制定目的、自我评审、适应范围、人员职责、政策贯彻、制裁等七个方面来确保英国 e-Science 项目建设的正常开展。

在该政策的制定目的中, 制定者明确指出, 制定本政策的目的是, 在英国实施 e-Science 项目的大环境之下, 以最完善的实践行动来保证信息安全。

在实施部分, 制定者指出: 支助项目人员接受相关的培训, 以便采用安全方法并在项目一开始就已设计出安全特征; 在项目早期阶段就得进行详细的威胁与风险分析; 提前制定一份详细的项目安全政策草案, 或者还得附上该政策的客观评论; 需要进行项目安全审查的特殊地点, 并提供资金让审查得到顺利进行; 与访问设备、数据集等有关特殊条件; 保持特殊安全技术在发展上与时俱进; 如果违背这些规定, 将会受到制裁。

尽管该政策有支持项目必须考虑安全问题的意图, 但是, 它也提出了相关制裁

措施。假如一个项目肆意或者是因为疏忽而让项目本身或者其他方面的安全面临危险，就有可能受到制裁。即，如果一个项目未能遵照它在获准通知书上所宣称的安全政策，或者未能采用与项目所面临的风险相关的安全方法，该项目就会受制裁。通常来说，制裁会跟不履行的性质的相匹配。制裁的尺度包括，从拒绝访问共享的数字科研设备到撤消已经授予给项目的资源。

(6) 英国 e-Science 评估机制及评估指标的特征分析

从 2003 年起，在英国 e-Science 指导委员会的监督指导下，英国国家 e-Science 中心开始对已经完成的或者正在进行的 e-Science 项目实施评估，并出台了若干份评估报告。

为了反映最新发展状况、总结经验以便指导以后 e-Science 建设与研究的开展，英国国家 e-Science 中心率先于 2004 年推出英国 e-Science 核心计划的建设进度报告，到了 2005 年，英国 e-Science 核心计划和英国研究理事会总会共同合作，推出 2004 年各研究理事会 e-Science 项目的评估报告。

在这份报告中，英国 e-Science 核心计划以及各个研究理事会对于 e-Science 的建设情况所使用的评估指标都不尽相同（详见附录 I）。尽管一些指标会被多个研究理事会共同用于 e-Science 建设状况的评估，但是，各研究理事会所使用的指标都与自身所承担的项目涉及的学科领域密切相关，同时也与各自的建设现状密切结合，具有很明显的个性化。

① 英国 e-Science 核心计划、七大研究理事会 e-Science 评估指标数量比较

纵观英国 e-Science 核心计划（CORE）与七大研究理事会 e-Science 评估指标可知（详见附录 J），若不比较各自使用指标的重叠性，则八大 e-Science 建设与研究主体到目前为止所使用的评估指标已经达到 122 项。而设置的指标数量最多的是工程与物理科学研究理事会（EPSRC），达到了 22 项，是粒子物理研究理事会所设立的评估指标数的两倍。这一方面反映了工程与物理科学研究理事会对 e-Science 建设状况在评估上的全面、深入与细致，更是可以揭示出，由于代表其他六个研究理事会管理英国 e-Science 核心计划，因此，在 e-Science 建设上，它具有更多的优势，并最终使得自身的 e-Science 建设方面走在所有研究理事会的前列。

另外，从指标的选取角度来看，各大 e-Science 建设与研究主体倾向于使用定量指标来评估各自 e-Science 建设与研究现状，表 10-5 充分说明了这一点。而既可以从定量的角度，也可以从定性的角度加以描述的评估指标，则在各大研究与建设主体所设置的评估体系中较少出现。

不过，对于定性评估指标，尽管各大 e-Science 研究与建设主体均有所设置，但是，在定性描述方面，大多都比较简略，要么泛泛而谈，要么就是一笔带过。例如，关于“基于网格的解决方案”的评价，既有从质量与影响等定性角度进行评估，也有从数量这一定量角度加以评估。不过，对于采取定性角度进行评估的研究理事会，却往往在这一栏目采用“为时尚早，不便评论”等表达，如医学研究理事会、经济与社会研究理事会；而采取定量角度进行评估的研究理事会则能在这一评价指标下

给出具体数量，如生物技术与生物科学研究理事会、中心实验室研究理事会。出现上述的主要原因在于，网格在 e-Science 中的应用为时尚早，投入 e-Science 计划中进行建设也只不过两三年的时间，其深远影响仍有待时日，因而现在谈论其影响与质量确实不够成熟。由此可见，到 2005 年底，英国各大 e-Science 建设与研究主体在评估 e-Science 建设进展上主要采取定量分析的方法。

表 10-5 CORE、七大研究理事会 e-Science 评估指标数量比较

	CORE	BBSRC	CCLRC	EPSRC	ESRC	MRC	NERC	PPARC
指标总数	17	16	11	22	15	12	18	11
定量指标数量	8	10	3	16	10	7	12	8
定性指数数量	8	5	5	4	5	5	5	3
定性定量结合	1	1	3	2			1	

② 英国 e-Science 核心计划、七大研究理事会 e-Science 评估指标差异性分析
通过将八大 e-Science 研究与建设主体所建立的 e-Science 评估指标体系实施差异性分析（gap analysis）可知，经过去重与归并。实际上，英国目前在 e-Science 建设与研究现状的评估方面，主要选取的评估角度有 46 项。

正如附录 J 所示，标准建设、国际合作与交流、e-Science 专业知识的培训与推广、科研成果是英国 e-Science 各大建设与研究主体普遍使用的评估指标。事实上，这也表明，作为一种全新的科研协作模式，e-Science 营造与终将实现的全球范围内科研资源的共建共享共知，在此过程中，标准与协议至关重要，而重视国际合作，注意经验交流，则是让科研人员提早感受到 e-Science 环境下合作的崭新性、便利性、全面性与实时性。

为了推进 e-Science 的进一步发展，如何让更多的人掌握 e-Science 专业知识，以进一步推动 e-Science 建设，普及 e-Science，显得至关重要。将 e-Science 专业知识的培训与推广作为当前 e-Science 项目建设进展的一大评估指标，充分显示了大多数研究理事会对于 e-Science 普及与推广的重视，也从一个侧面反映了当前 e-Science 建设与研究的时代特征。

除此之外，承担的项目数量、参与的 e-Science 专题会议数、投资额、成果的应用领域与范围也是当前各大 e-Science 研究与建设主体较为关注的问题。随着 e-Science 建设与研究的深入，各大建设与研究主体将对自身的 e-Science 建设与研究评估指标体系做出相应调整将是一种必然。

在评估开展的过程中，除了有网格工程工作组直接参与到与中间件相关的项目评估中，包括各研究理事会在内，不仅对自身承担的项目进行评估并形成评估报告提交英国科技办公室审核，而且还参与了自身没有直接参加的其他项目评估活动中。在进行项目评估的过程中，研究理事会的各学科委员分别会按领域受理项目评估，参与评估的科学家以“评判员”的身份给每份项目报告提出评估意见。纵观英国 e-Science 项目的评估过程，不难发现，及时、公正、公开、公平是它的出色之处。

2. 欧盟e-Science管理运行机制

欧盟框架计划是实现欧洲研究领域（ERA）目标的一个重要的研究资助计划，欧洲的科研活动即是在框架计划下以每连续四年为一个阶段进行资助和管理。欧盟FP6下设定了七大优先研究的主题领域，其中“信息社会技术”（IST）就是七大主题之一，该主题下成立了一个咨询专家组（ISTAG），该专家组主要建议欧盟在IST优先领域下研究活动的总体战略，并负责建议和实施欧盟在信息通信技术（ICT）研究中的资助政策。

欧盟的e-Science研究项目的资助管理统一于欧盟框架计划下，欧盟已经先后在第五框架计划和第六框架计划下总共资助了40个项目，总资助额达1.1亿欧元。e-Science项目的资助管理主要由欧盟委员会下的“信息社会和媒体总理事会”负责。

（1）信息社会和媒体总理事会

信息社会和媒体总理事会（Information Society and Media Directorate-General）是欧盟委员会下的一个信息通信技术（ICT）委员会。该总理事会的使命是支持欧洲范围内信息通信技术的开发和利用。该总理事会下共有10个分理事会，其中e-Science相关的研究活动主要在“新出现的技术和基础设施”主题下，由分理事会F负责组织管理。F理事会设有主任一名，该理事会的研究资助活动主要集中在下面四个领域：① 未来新出现的技术；② 未来一代网格技术；③ 计算机和通信研究基础设施；④ 网络知识经济中新的工作环境。在F理事会下e-Science研究的资助管理又按照项目性质分别由两个工作单元管理，其中与网格技术相关的e-Science项目由F2工作单元负责资助管理；而与研究基础设施相关的e-Science项目则由F3工作单元负责资助管理。此外，除F理事会之外的其他理事会也会在其特定的主题领域下资助并管理一些e-Science项目。

（2）eInfrastructure 审议工作组

eInfrastructure 审议工作组（eInfrastructure Reflection Group, eIRG）是欧洲范围内协调e-Science基础设施研究的一个重要机构。该机构的主要目标是在政治、咨询和监管层面上支持使欧洲范围内电子资源有效而方便共享的政策和管理框架的创立。eIRG的具体目标包括：① 识别促使泛欧洲e-Science实现的基础设施、服务和资源需求；② 为欧洲地区性和国家性的网格计划推荐资源共享的政策指南，以及为国际政策制定作贡献；③ 首先聚焦于e-Science应用的用户，其次关注更广泛领域内的应用（包括e-Government、e-Health等）；④ 在相关利益团体中宣传网格技术；⑤ 关注网格部署中的管理问题；⑥ 学习国家研究与教育网（NREN）团体的经验。该机构主要由来自欧洲各国教育部的官方代表组成。

（3）典型e-Science项目的管理运行

① DataGrid项目的管理运行实例

DataGrid项目是欧盟资助的第五框架计划下的网格项目，该项目由欧洲原子能

研究组织 CERN 牵头，共有 15 个机构参与。DataGrid 项目由信息社会和媒体总理事会 F 理事会下的 F3 工作单元负责监管。该项目设有一名项目经理，并有两个委员会协助项目管理。一个是项目管理委员会（Project Management Board），另一个是项目技术委员会（Project Technical Board）。项目管理委员会主要负责管理项目的 15 个参与机构，而项目技术委员会则负责管理项目的工作组（Work Package）。DataGrid 项目总共有 12 个工作组，每一个工作组设有一名工作组经理。此外，在项目经理下设有项目办公室和质量保障部，项目经理下还有一个技术协调人负责协调项目的 12 个工作组（图 10-34）。

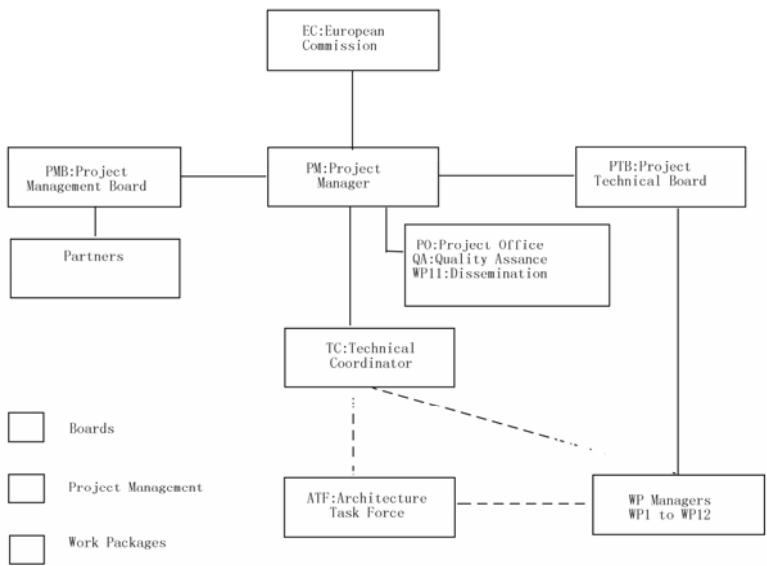


图 10-34 DataGrid 项目组织结构图

② CoreGrid 项目的管理运行实例

CoreGrid 项目是欧盟资助的第六框架计划下的网格项目，该项目由欧洲信息和数学委员会（ERCIM）牵头，总共有 42 个结构参与。CoreGrid 项目由信息社会和媒体总理事会 F 理事会下的 F2 工作单元负责监管。CoreGrid 项目有一个整合监管委员会（Integration Monitoring Committee）。在该委员会之上有两个咨询委员会，一个是科学咨询委员会（Scientific Advisory Board，SAB），另一个是工业咨询委员会（Industrial Advisory Board，IAB）。SAB 有 6 个成员，其中 3 个来自内部科学家，3 个来自外部。SAB 的主要目标之一是作为同行评议理事会，评估项目工作进展和项目成果的总体科学质量。IAB 有 15 个来自工业界的成员，它的主要目标是促使项目与工业界的沟通，确保该项目对欧洲工业界产生重要影响。CoreGrid 项目的管理运行主要由项目执行委员会（Executive Committee）负责，该委员会设有两个协调负责人，分别是行政与财务协调负责人和科学协调负责人。并且有一个项目参与机构组成的成员全体大会（Members General Assembly）协助执行委员会的管理工作。

CoreGrid 项目的主要活动被划分为 6 个组，称为 CoreGrid 虚拟研究所（Virtual Institute）。这 6 个虚拟研究所分别是知识与数据管理、系统架构、日程安排、编程模型、网格信息与监测、问题解决环境工具与网格系统。在整个项目中还有两个工作组，其中一个项目活动整合（Integrated Activities）工作组，它主要负责项目内部各工作组之间的协调和整合；另一个是成果推广（Spreading Excellence）工作组，它主要负责项目成果在欧洲网格研究团体中的推广。CoreGrid 项目组织结构中最后一个工作组被称为合作网关（Collaboration Gateway），它的主要任务是管理 CoreGrid 与欧盟内其他相关网格项目之间的合作和整合（图 10-35）。

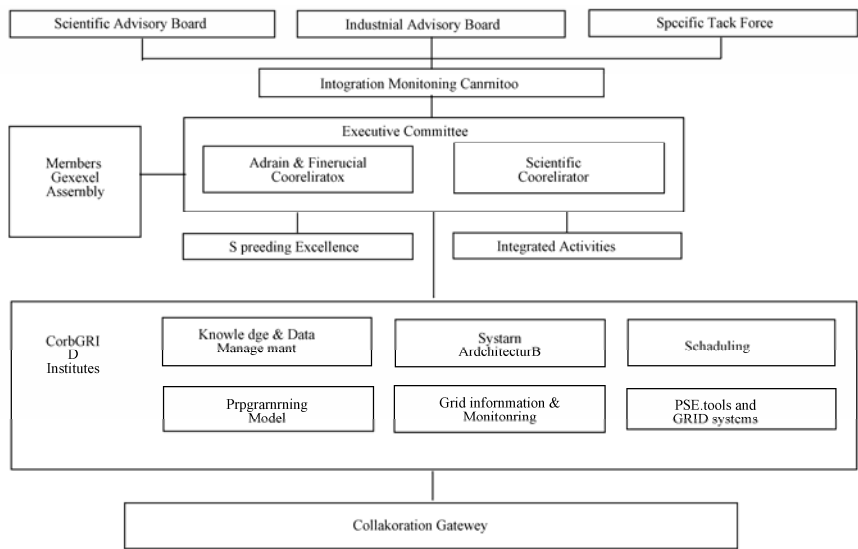


图 10-35 CoreGrid 项目组织结构图

总之，欧洲 e-Science 的项目运行整合在整个欧盟框架计划中，并由欧盟委员会下的信息社会和媒体总理事会及相关的分理事会统一监管。在第五框架计划下，总共启动了三类 e-Science 项目，分别是研究技术开发（RTD）类、辅助设施类和主题网络类，其中以 RTD 类项目运行的最多。在第六框架计划下，总共运行了四类项目，分别是整合研究项目、卓越网络项目、特定目标研究项目和特定支持行动项目。虽然框架计划从第五过渡到第六，但是 e-Science 项目的管理运行机构并没有变化，这种稳定的管理运行机制保证了欧洲 e-Science 研究的延续性。

10.8.2 以美国为代表的分散式管理运行模式

e-Science 建设的分散式管理运行模式是指在 e-Science 建设过程中，缺乏一个全国性或地区性的统一规划，一国或一地区的 e-Science 项目没有统一的管理运行机构，本国或本地区的 e-Science 研究分散于各个研究机构中，并由各个研究资助机构分别管理项目的运行。

分散式管理运行模式以美国为代表。美国的 e-Science 研究没有统一的管理运行机构，取而代之的是各资助机构通过其提供的各种资助计划进行的分散管理。这种分散式的管理运行机制的一大优点是对于具体的应用需求有了更强的针对性，然而其不便于管理的缺点也显而易见的。这较为突出地表现为项目资金来源的多元化——同一个项目从多个资助机构的多个资助计划获得资助，有的项目甚至通过同一个资助机构的多个资助计划获得资助，尽管这样造成的结果是对于单个项目有较为充裕的资金来源以保证研究的顺利进行，但对于 e-Science 研究的整体评估和投资的宏观管理和决策则会产生一定的误导。

10.9 政府是e-Science建设的主要投资者

目前，不管是英国、美国，还是欧盟，在 e-Science 的投资方面，政府均发挥着主导作用。在英国，2000 年 11 月，英国政府宣布投资 2.5 亿英镑用于英国的 e-Science 建设，标志着英国大规模开展 e-Science 建设的开始。在政府提供的这笔款项中，7400 万英镑由政府科技办公室直接分配给英国的 7 个研究理事会，用于其进行各自与 e-Science 相关的研究。在美国，承担 e-Science 相关项目的主要机构有 DOE、NSF、NASA 等，得益于美国政府对网格技术建设的重视，其投资也由政府承担。NSF 于 2001 年启动的 NMI 计划共获得美国政府 4000 万美元的拨款，同年由 NSF 启动的 ITR 计划也获得美国政府约 1.3 亿美元的支持。而在此之前，美国国会拨款近 6 亿资助 DOE 启动 DOE SciDAC 计划，用于建立科学计算、硬件框架，以开发大规模计算在先进科学发现和工程应用方面的潜能。

与此相比，欧盟委员会在其启动的第五框架计划、第六框架计划中，先后投入 1.1 亿欧元，充分体现出欧盟委员会在财政方面对 e-Science 相关建设项目的大力支持。此外，欧洲各国政府也是建设国家性 e-Science 计划的主要投资者。1998 年法国启动了 RNRT 计划，该计划由法国工业和电信部联合资助。在 1999—2002 年四年间该计划总共资助了 200 多个项目，政府总资助额高达 1.85 亿欧元。2000 年法国研究部和工业部共同资助发起了 RNTL (Réseau National des Technologies Logicielles) 计划。在 2000—2002 年三年间共有 120 个项目被资助，总资助额达 1.91 亿欧元。2001 年启动的 ACIGRID 计划是法国 e-Science 发展中的重要一步，该计划由 FNS 单独资助，每年资助 15 个项目，平均每年的资助预算是 300 万欧元。2004 年 6 月启动的 VIOLA 项目由德国 BMBF 资助，资助额为 1000 万欧元。2004 年 9 月德国联邦教育与研究部 (BMBF) 资助的 eSciDoc 项目，总资助额为 610 万欧元。2005 年 9 月德国启动了国家 e-Science 框架计划——D-Grid 计划。该计划是目前德国最大的 e-Science 计划，由德国 BMBF 资助，5 年总资助额达 1 亿欧元。2002 年 9 月被正式批准的意大利国家网格研究项目 GRID.IT 由意大利教育大学研究部 (MIUR) 资助，总资助额为 810 万欧元。荷兰面向 e-Science 的虚拟实验室 (VL-e) 项目政府资助额度为 2000 万欧元。所有这些足以说明各国政府在 e-Science 投资建设的投资中起到主导作用。

第 11 章 中国e-Science规划与建设分析

11.1 中国e-Science规划与建设的主要特征

1. 中国已经具有e-Science规划雏形

与英国相比, 尽管中国仍未在全国范围内开展大规模的 e-Science 规划与建设, 但是, 中国已经做好这方面的准备、蓄势待发。2004 年, 在英国开始实施 e-Science 核心计划第二阶段的同一年, 科技部、国家发改委、教育部、财政部则联合制定并出台了《2004—2010 年国家科技基础条件平台建设纲要》(2004) (下文简称《纲要》), 明确提出国家科技基础条件平台的框架, 及今后一段时间内全面推进国家科技基础条件平台建设的指导思想和原则、目标和任务、建设重点和保障措施。

在建设目标上, 《纲要》指出, 到 2010 年, 初步建成适应科技创新需求和科技发展需要的科技基础条件支撑体系, 以共享机制为核心的管理制度, 与平台建设和发展相适应的专业化人才队伍和研究服务机构。为最终形成布局合理、功能完善、体系健全、共享高效的国家科技基础条件平台奠定基础。

为保证上述目标的实现, 《纲要》指出, 国家科技基础条件平台建设的三项重要任务:

(1) 构建和完善物质与信息保障系统。制定科学、合理、统一的技术标准和规范, 研究开发相关技术, 对现有的大型科学仪器、设备、设施、科学数据、科技文献、自然资源资源等进行整合、重组和优化, 充分利用国际资源, 加快实现资源的信息化、网络化, 建立适当集中与适度分布相结合的资源配置格局。

(2) 建立以共享为核心的制度体系。制定、公布《科技资源管理法》, 加快推进修改、制定一系列配套的法律、法规、规章和标准, 明确各相关主体的责任、权利和义务, 建立和完善激励机制和评估监测机制, 推进管理方式创新, 创造公共资源公平使用的法制环境。

(3) 培育专业化的人才队伍和机构。深化科研机构人事制度改革, 完善评价体系, 建立人才凝聚机制, 培育、形成一支专门从事科技基础条件管理与技术支撑的人才队伍。

在此基础上, 《纲要》希望通过五个基础平台的建设完成上述任务。这五个基础平台是: 研究实验基地和大型科学仪器、设备共享平台; 自然资源资源共享平台; 科学数据共享平台; 科技文献共享平台; 成果转化公共服务平台。

从 e-Science 理念的本质上分析, e-Science 是关于关键科学领域的全球协作的技

术，以及与之相适应的下一代体系结构。e-Science 的这种分布式、全球大规模合作的科学研究方式将有助于科研人员获得海量的数据、大量的计算资源以及高性能可视环境等。e-Science 对于非常昂贵和稀缺的科研设备（如同步加速器、卫星、哈勃望远镜等）的共享具有非常重要的作用。通过 e-Science 技术，科研人员可有效地访问和管理在地理上分布的数据、计算和网络资源，并及时地分析这些海量数据（如进行数据挖掘和知识发现等）。

因此，无论是从《纲要》所指出的国家科技基础条件平台建设的建设目标、建设任务，还是从实现途径来看，通过加强 e-Science 的建设，无疑是建设强有力国家科技基础条件平台的重要实现途径。从这一点分析，《纲要》无疑是我国建设 e-Science 的重要指导性文件，显示了我国已经具备了 e-Science 建设的规划雏形。

2. 中国以网格为先导，为e-Science发展打下基础

e-Science 和网格技术紧密相关。一般认为，e-Science 的需求只有通过网格技术才能实现。目前，纵观我国 e-Science 的发展现状，尽管我国仍未开展大规模的 e-Science 建设，但是，自 1999 年以来，我国已投入大量资金、人力对 e-Science 的核心支撑技术——网格技术进行研究开发并已取得不错的成果，体现出了我国在 e-Science 方面的建设特色——以网格为先导。

早在 1999 年，中国就已启动了全国性的网格建设项目——国家高性能计算环境（HPCE）项目，同年，在教育部的支持下，李三立院士带领清华大学网格研究组进行了先进计算基础设施 ACI（Advanced Computational Infrastructure）的研究。该项目将分布于北京和上海的两台自主研制的超级计算机连接成为聚合计算能力达到 4500 亿次的网络计算平台，并提供网页访问界面，让分布于各地的用户能够在家中使用的超级计算能力。随后，“国家高性能计算环境（NHPCE）”、863 空间信息网格项目、知识网格项目等项目纷纷投入建设，到了 2002 年，科技部召开了“网格战略研讨会”，确认将网格的研究和应用列为“863 计划”的一个专项，随即成立了专项专家组，并拨出 3 亿专款用于网格的研究与建设。自此以后，中国便掀起了一股网格研究与建设热。据不完全统计，截止到 2005 年 12 月底，中国已经开展过或正在开展的网格研究与建设项目已达到 30 个。

在网格技术研发蓬勃发展的基础上，国内各界开始探讨中国 e-Science 的建设问题并采取相应措施。2002 年 6 月，英国国家 e-Science 研究计划重大项目首席科学家 Tony Hey 教授、Stephen G Oliver 教授、中国国家 863 计划网格重大专项总体组组长钱德沛教授、中国国家 863 计划生物和现代农业领域生物信息技术主题专家组长李亦学教授分别代表中英两国科学家就双方在生物信息学、药物设计、生物多样性、结构基因组学、生物网格建设等上述五个方面的 e-Science 合作研究达成协议。至此，中国 e-Science 开启了以网格技术推动 e-Science 建设的新篇章。

目前，借助网格技术，中科院已经在 e-Science 方面做了一些基础性工作。2001

年,中科院就开始以科学数据网格作为 e-Science 的一项重要基础设施进行初步建设,并希望通过该网格项目的建设,连接中国科学院分布在全国的四十多个研究所,通过网格计算的相关技术,特别是数据网格技术,实现对科学数据库中大量分布式异构数据资源的有效共享,并在此基础上开发对科学研究有直接实用意义的应用系统。

除此之外,初步以上海高校网格、计算科学 E-研究院、免疫学 E-研究院、模式生物 E-研、都市文化 E-研究院、社会学 E-研究院等六大网格建设项目为支撑的上海 E-研究院(e-Institute)也于 2002 年投入建设。该项目由上海市人民政府投资 2 亿多人民币启动,是一个以信息网络为平台、全新的具有可变性、超大容量的虚拟研究机构。它不需要一般研究院那样的庞大人事组织机构,仅需通过约定和网格就可以与全市、全国乃至全世界的研究人员对重大课题、重大攻关项目进行研究上的合作,还可以利用网格技术在世界范围内共同进行研究生的培养和教育。从 e-Science 的建设理念分析,该项目堪称我国当前 e-Science 建设的大手笔。

中国虽然未有明确实施大规模的 e-Science 建设,但是,以五大网格为底层支撑,以国家科技基础条件平台建设为规划指导,却让人初步感受到中国 e-Science 的建设雏形。

1. 中国发展e-Science的需求与机遇

各行各业都在大规模建设网络基础设施,大量采购服务器、计算机、软件;大力开发应用系统,传统的业务正在向计算机化、网络化方向发展。

2. 中国e-Science的发展契机——五大网格的支撑

网格支撑是中国 e-Science 发展的契机,其作用包括以下 3 方面。

(1) 网格技术就是应对这些问题而出现的新技术,将为国家信息化建设,经济和社会可持续发展发挥重要的作用。

(2) 网格支持资源共享、协同工作,将成为新一代信息基础设施。

(3) 利用中国已经建成的五大应用网格,严重阻碍着我国信息化进程的矛盾和问题将得到解决。

总之,对于 e-Science 建设来说,网格技术带来的三大利益表现为:降低而不是提高应用人才的门槛;降低而不是提高应用开发难度;降低而不是提高应用运行成本。利用我国在网格方面的建设成果,将为我国开展在大规模的 e-Science 建设奠定坚实的基础。

3. 集中采用两种模式完善e-Science体系结构

e-Science 目前在国内的发展还没有一个完全统一和固定的发展模式,已有的项目按照自身发展需求采取相应的发展模式,有的是各层之间并行推进,有的是从底层开始,层次递进。

(1) 网络中间件层与应用层的并行推进

中国国家网格以网络中间件为建设重点, 兼顾应用层的建设。目前, 网络中间件层已经能够提供网络监控、作业管理、用户管理、信息管理、记账管理、安全、Web Portal 等功能模块, 在应用层的建设方面, 中国国家网格建立了 11 个应用网格, 即气象网格、资源环境网格、航空制造网格、科学数据网格、新药研发网格、生物信息网格、仿真应用网格、教育网格、城市交通信息服务网格、国家地质调查网格、森林资源与林业生态工程网格等。

(2) 层次递进型——中国科育科研网格

以中国教育科研网为例, 前期工作重点是将建设与布置网格基础设施, 包括各节点的硬件设施配置建设, 网格操作系统的研发, 人员的培训; 在前期网格基础设施建设的基础上, 后期着力于建设图像处理网格、生物信息学网格、海量网格、计算流体力学网格以及远程教育网格。

依照本书编者提出的 e-Science 建设五层架构理论, 那么, 结合中国教育科研网格的建设情况, 其建设模式呈现为以底层为基础、层次推进。即先从网格基础设施、网络中间件等底层基础进行建设, 再逐步加强应用层面的建设。

4. 注重合作

中国教育科研网格与各大公司合作建设 China Grid 的公共平台。不过, 这个公共平台目前在国际上并没有统一的标准, 也没有很好地完整实现。China Grid 正在与包括英特尔、惠普、IBM 在内的公司展开关于平台方面的探讨和研究, 并准备开展合作。

中国国家网格(CNGrid)在建设的过程中, 就密切注意与国外的 e-Science 建设单位加强合作。例如, 项目组于 2002 年 6 月和 2003 年 1 月两次参加中英生物信息方面的会议, 到了 2003 年 10 月, 项目组再次访问英国 e-Science 有关部门, 探讨与 e-Science 互联和进一步合作。

在建设网格的过程中, 国内有关项目组相当重视与国外相关单位共同成立合作机构。根据中英两国在 e-Science 和网络方面的部长级合作备忘录精神, 中方通过建立开放中间件基础架构研究所(中国)(OM II -China)与英方的 OMII-UK 开展合作, 其主要任务是联合研究、开发开放的网格中间件以及可以被其他网格软件平台集成的面向服务的软件构件, 研究开发软件测试和软件质量保证的方法, 以保证其开发的软件符合 OM II 的质量标准。OM II -China 是由多所中国的大学和研究所联合组成的, 将在开发符合 OM II 标准的网格软件的国际合作中作出重要贡献。

5. 初步形成规范的交流机制

在开展网格建设的过程中, 中国各大网格建设主体均充分认识到国际交流的重要性, 通过多种途径与国际同行开展交流。目前, 这种交流途径主要有: 召开国际

会议、参与国际竞赛（如中国国家网格）、培训（如教育科研网格）。

另外，部分网格项目组已经形成统一的交流制度。例如，中国国家网格制定实施了季报制度与不定期汇报制度：各节点必须按季度向中心汇报各自的最新建设进展；针对中心运行中出现的技术问题，采取不定期汇报的方式，及时描述故障现象、分析故障原因，上报事故处理情况。

6. 初步的国际影响力

在武汉召开的网格计算会议上有包括 GGF（全球网格论坛）主席、两位网格之父以及英国网格计划（该计划投资 2 亿英镑把整个英国国家的资源集合在一起形成一个大平台）负责人等在内的权威参会甚至做主题演讲，这也可以看出国外学术界对我国网格应用研究的重视。

另外，在 2005 年 10 月 14 日刚刚结束的法国 GRID@WORK 大会上，中国国家网格与其他国际网格项目密切合作，成功为大会举办的第二届 Plugtests 国际网格大赛提供了网格设施服务。而在网格计算能力的支持与服务方面，中国国家网格排名第二，成为迄今中国最大规模网格设施参与实质性国际合作的领头羊，这次国际合作锻炼了国家网格的组织协调与国际合作能力，大大提高了中国国家网格的国际知名度，引起国际同行的广泛关注。

7. 注重引进国外先进的技术

China Grid 的成员在今年召开了一次连续四天的专家组会议，并且请国际上的著名公司来培训，学习这些公司最先进的技术。例如，惠普的科学家就谈到了欧洲的大规模碰撞的计算网格是怎么做的，并专门从英国派人来介绍网格里面的安全设计。紧接着，China Grid 小组通过召开专家会进行深入的探讨，并一步步地把最先进的技术引进来。

8. 推进技术标准化

ChinaGrid 公共支撑平台（CGSP）是 ChinaGrid 为各种网格应用提供基础支撑的网格核心中间件。CGSP 是国际上第一个遵循 OGSA 架构，参照 WSRF 规范实现的网格中间件。

除此之外，NSFCGrid 建设规划下的 CROWN 项目在开展的过程中，非常注重与推进技术标准化。例如，向 IETF63 届大会提交标准化草案；提出 ETF gar 封装标准；接受 UK e-Science ETF 对 CROWN 中间件系统进行相关评估工作；加入国际标准组织 OMG，参与 OMG 组织的 5 项技术标准制订进程。

11.2 中国e-Science规划与建设与国外的差距分析

尽管我国的 e-Science 研究与建设已经在神州大地轰轰烈烈地展开，并且经过相关人员多年的努力也取得了可喜的成绩，但是，纵观世界各国、地区在发展 e-Science 方面取得了成就，我国与发达国家之间仍存在不小差距，中国的 e-Science 进程要想真正成为推动我国科学研究的有力工具，仍然需要我们坚持不懈的辛勤耕耘。下面就具体从 e-Science 的网格基础设施层、网格中间件层、应用开发环境与工具层以及具体应用层来分析我国在 e-Science 规划与建设中与国外发达国家之间的差距。

11.2.1 从网格基础设施层分析我国的发展差距

我国在网格基础设施层的主要建设项目包括：中国国家网格、中国教育科研网、上海网格。纵观中国的网格基础设施建设现状可以发现，我国在网格基础设施的建设方面已初具规模，不仅连接了大量科研院校的资源 and 设备，形成了大规模的计算能力和数据存储能力，而且在现有网格基础设施的基础上建立起了一批应用网格，实现了将分布异构的资源应用到对具体科研问题的解决方面的支持。然而，与世界发达国家的网格基础设施建设相比，我国的网格基础设施建设还有待加强。首先，在连接的资源方面，欧洲的 EGEE 项目已经将 27 个国家的 71 个机构组织的资源连接到了一起，并拥有超过 20000 个 CPU 的合计计算机能力，这是世界上集成的最庞大的网格基础设施，而美国的 DOE Science Grid 项目也已经连接了 35 个站点、3500 台 CPU、4 个国家实验室和数量众多的大学。另外，在对网格基础设施的使用方面，EGEE 上面运行着高能物理领域的 LHC 项目、美国的 Babar 项目、CDF 项目、德国的 ZEUS 项目、生物医学领域的 WISDOM 药物发现、Mammogrid、GATE、GridGRAMM 等 14 个项目，此外，还包括数字图书馆领域的 GRACE 项目、计算化学领域的 CompChem 项目等 12 个应用项目；而 DOE Science Grid 上则运行着 HEP、LIGO、SDSS、GENO、CS、iVDGL、GriPhyN、PPDG、US LHC 等大型的应用网格。可见无论是在网格基础设施所整合的资源方面，还是基于网格基础设施所进行的具体科研应用方面，我国都有待进一步提高，以跻身世界领先的行列。

11.2.2 从网格中间件层分析我国的发展差距

1. 网格操作系统

当前，Globus Toolkits 是全球最通用的网格操作系统，这个由全球网格论坛（GGF）下属 Globus 项目组成员联合开发的 Globus Toolkit 标准工具包，已被公认为当前建立网格系统和开发网格软件事实上的参考标准。Globus 项目是国际上与网

格计算相关的最有影响力的项目。但是值得一提的是，中国的两大网格都自行开发了能够与 GT 相匹敌的网格操作系统(GOS)，分别是 CNGRID 的 VEGA 织女星 GOS 和 ChinaGrid 的 CGSP (ChinaGrid Supporting Platform)。下面就初步对这 3 个 GOS 进行比较分析。

(1) Globus Toolkit

由 Globus Alliance 开发的 Globus Toolkit 是网格技术框架的最优秀的实现者与拓展者，为网格项目的建设者提供最有力的技术支持，而且其所有源代码都可以免费获得。Globus Toolkit 实际是问题解决方案的集合，它包括那些在实施网格系统和应用的过程中被证明有用的组件，这些组件被通用化，以使其能够在更广泛的应用中被加以利用。但 Globus Toolkit 并不是要解决所有问题，真正全面的解决方案需要多个组件的组合使用。

Globus Toolkit 从 1.0 版本到 4.0，一直都十分关注在异构平台间进行合作的研究，其为应用开发者提供实现对异构系统组件交互的标准界面工具以消弱异构性。这些工具非常适合 OGSA 框架中资源层及其以下的层次，而且相信 Globus 会提供更加“垂直化的解决方案”（解决不同应用类型中特定问题的工具，如数据复制、元数据管理和授权等问题），在现有资源层的基础上以“收藏层”服务提供更强大的功能。

Globus Toolkit 一直被设计和实施来利用和鼓励使用现有来自于如 IETF, W3C, OASIS 和 GGF 的标准。其所开发的工具都是使用这些标准，而不是使用新的实现机制，例如：SSL/TLS v1 (from OpenSSL) (IETF)、LDAP v3 (from OpenLDAP) (IETF)、X.509 Proxy Certificates (IETF)、SOAP (W3C)、HTTP (W3C)、GridFTP v1.0 (GGF)、OGSI v1.0 (GGF) 等。

同时 Globus Toolkit 也在使用这些组织新推出的标准，并提供代码和收集使用反馈。例如，WSRF (GGF, OASIS)，DAI (GGF)，WS-Agreement, WSDL 2.0 (W3C)，WSDM, SAML (OASIS) 和 XACML (OASIS)。

为减少根据定制要求组装组件所遇到的问题，Globus Toolkit 试图提供一套标准的系统组件以支持广泛的特定应用的定制要求和用户界面要求，而不是只提供一个唯一的架构。不过需要注意的是，Globus Toolkit 的组件并不能提供一个完全的解决方案，而是要根据具体要求进行调整和适应。这种“搭积木”的方法在实践中，已经被证明是一种非常有效的解决方法。

(2) VEGA 织女星操作系统

织女星网格体系结构的基本思想是把网格看成一台虚拟的超级计算机系统。与现有的计算机系统类似，织女星网格也将包括硬件、系统软件和应用三个组成部分，而织女星网格操作系统是网格硬件资源的管理者。

织女星网格操作系统主要分为 4 个功能模块，分别是 GOS (Grid Operation System) 应用、GOS 系统服务、GOS 核心服务、GOS 主环境，如图 11-1 所示。

实现的基本网格服务包括资源管理和数据管理两方面。资源管理：织女星网格操作系统实现对全网格计算资源的统一管理，隐藏计算资源的异构性、动态性和分

布性，提供可靠的资源使用方式；数据管理：在现有的计算机系统中，存储系统负责数据的管理。存储系统分为两个层次，即存储设备和文件系统。存储设备完成对数据的物理保存，而文件系统负责对这些数据进行逻辑上的组织。

另外，在织女星网格的研究中，提出了基于“网格计算协议”（Grid, Computing, Protocol, GCP）的网格互联系统。该协议由基于 TCP/IP，协议的资源路由协议和应用层次的网格计算表示协议所组成。

织女星网格操作系统借鉴了这种设计思想，对网格计算进行了抽象定义，提供一组具有特定含义的编程接口，用于开发上层织女星应用系统。

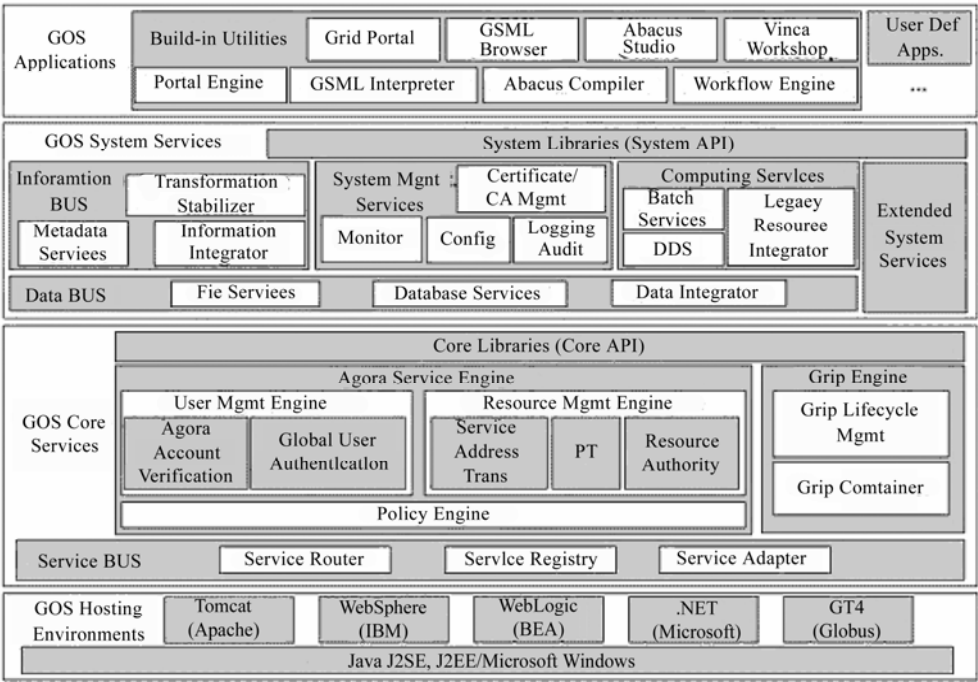


图 11-1 VEGA 体系结构

(3) CGSP

ChinaGrid 开发了一个为各种网格应用提供基础支撑的网格核心中间件——ChinaGrid 公共支撑平台（CGSP），并于 2005 年 1 月发布了 v1.0。很多专家对 CGSP 异常看好，认为完全与 Globus Toolkit4 有得一比，并受到了 Globus 项目的重视。

CGSP 是国际上第一个遵循 OGSA 架构，参照 WSRF 规范实现的网格中间件。它提供了一套完整的网格服务支撑平台，对教育和科研系统中的各种资源进行整合，屏蔽网格资源的异构性和动态性，为各种科学计算与工程研究提供高性能的、高可靠的、安全方便的透明网格服务，形成一套公共网格服务体系。目前已经有生物信息学、图像处理、计算力学等的应用在该平台上调试通过。

CGSP 是一组互相配合的软件组件，包含若干个可独立运行的软件，用于支持开发过程、运行过程、系统安装过程和系统管理的各个环节。其详细模块结构图如图 11-2 所示。

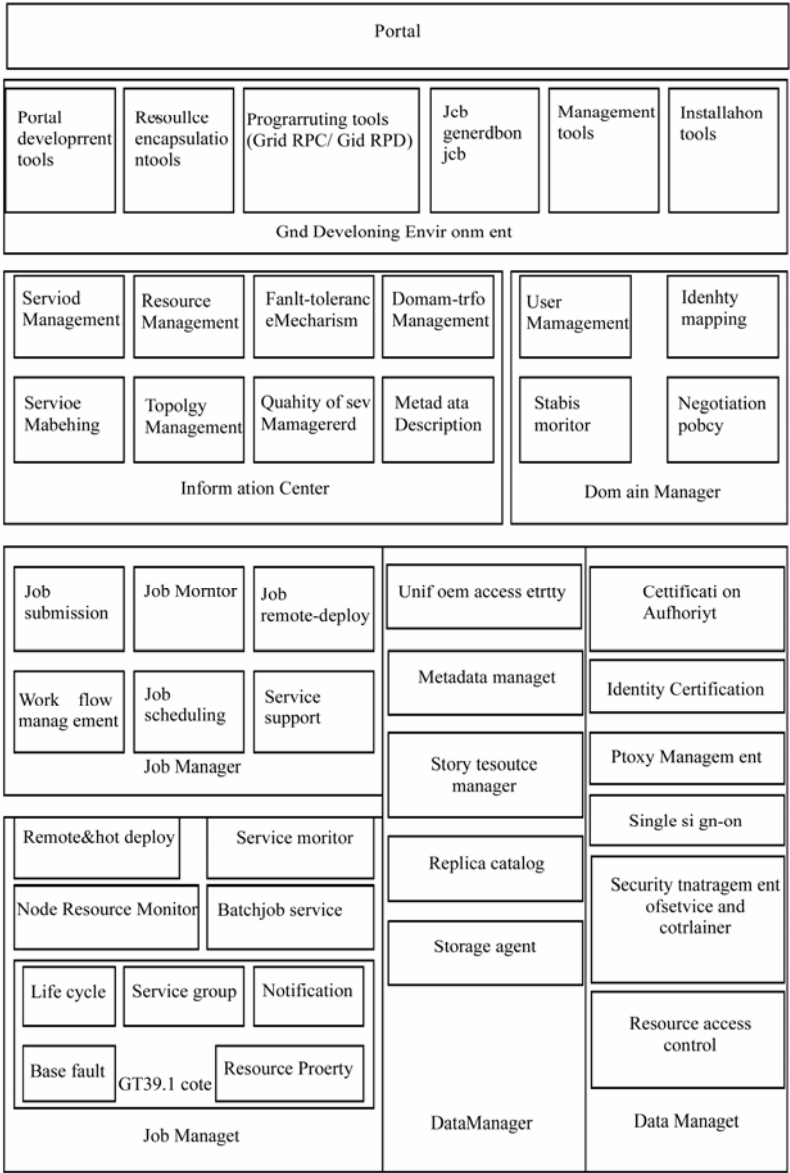


图 11-2 CGSP 框架

CGSP 共划分为六个功能模块：网格门户（Portal），作为 ChinaGrid 的网络服务展现方式。网格门户是最终用户使用网络的入口，并由此提供任务和获得结果；网

格开发环境，其主要功能是提供资源网格化封装的工具包和网格构建管理工具包，以及提供面向网格环境的编程模型，用于复杂网格作业的开发；信息中心，其主要功能是负责网格环境中各类资源信息的管理，实现一个全局的资源视图，提供网格信息服务，并实时更新网格资源信息；统一管理层，其主要功能是为网格环境中各类作业操作提供基础支持；网络安全，提供的功能包括用户身份的认证、资源和服务的授权、加密传输，以及用户身份到资源授权的映射等；计算节点层，真正提供网格服务的节点，可以是集群等大型计算设备，也可以是普通微机和工作站等。

通过对 3 个网格操作系统的分析，我们可以看出，中国的 2 个网格操作系统研究在整体的功能模块的部署和设计方面与 Globus 相比并不逊色，而且有异曲同工之妙。这一方面是由网格操作系统所要完成的基本功能决定的，另一方面也说明中国的网格技术人员在网格开发与建设的整体把握上，已较为完善和准确。但是，从目前的情况来说，中国的两个网格操作系统在对现有标准的支持以及对新标准的推动方面，似乎还没有像 Globus 那样彻底，当然，Globus 全球合作的背景是产生其重视标准的重要因素，不过为了将来能够顺利的与国际先进技术和成果的融合，中国在标准的使用和参与制定方面还应保持较高的关注力度。另外，Globus 整合了很多开源的优秀工具，不但减少了重复开发造成的人力、资源浪费，而且也达到了很好的实用效果，因此我国的网格操作系统研究也应该注意对已开发的开放源代码的软件产品的充分利用。

2. 网络中间件组件

根据英国 e-Science 中心的 3 层结构所设置的网格中间件角度出发，则中国的网格建设主要涉及如表 11-1 所示的模块研究和建设。

由表 11-1 可以看出，中国在网格层的建设总体上较为关注认证与授权、协同调度、数据访问服务、资源访问服务、安全与隐私、合作与远程设备仪器服务、代理和网格信息服务等组件的研究和使用；而对全局事件服务、全局请求、故障管理、服务质量的可靠性、网格监测和数据编目与溯源方面没有太多的网格建设项目涉及。归结起来，可以说中国的网格建设比较注重网格基本功能的实现，而对于网格质量的优化并没有给予更多的重视，当然这里指的是总体情况，对于个别网格项目在网格质量的优化方面还是做出了不少努力，比如科学数据网格和上海网格。另一方面，在美国和英国方面，对于网格层的研究情况如表 11-2 所示。

从表 11-2 中可以看出，英国在各组件的研究方面项目分布较为平均，而美国地区的网格研究则呈现出和中国类似的情况，即注重网格基本功能的实现，而对于提高网格服务质量方面的研究投入力度不是很大。不过，美国对于网格监测方面的研究还是比较重视的，并有 WS Core Monitoring Features、Globus Index Service、Globus Trigger Service、Ganglia Cluster Toolkit、Inca、MonALISA 等网格监测软件产品；而中国地区只有中国国家网格及旗下的中国气象应用网格和先进计算基础设施北京，上海试点工程 3 个项目涉及该方面的研究和建设。

表 11-1 中国网格层建设情况

网格层	组 件 名 称	项 目 数 量
	审核	4
	认证与授权	10
	代理	6
	合作与远程仪器服务	8
	协同调度	10
	数据访问服务	11
	数据编目与溯源	3
	故障管理	1
	全局事件服务	0
	网格信息服务	6
	全局请求	0
	监测	3
	服务质量的可靠性	2
	资源访问服务	11
	安全与隐私	12

表 11-2 美国、英国在网格层研究情况

网格层	组 件 名 称	美 国	英 国
	审核	2	3
	认证与授权	14	6
	代理	2	6
	合作与远程仪器服务	17	7
	协同调度	6	8
	数据访问服务	5	9
	数据分类与溯源	5	8
	错误管理	2	4
	全局事件服务	4	5
	网格信息服务	4	4
	全局请求	0	7
	监测	14	4
	服务质量的可靠性	1	7
	资源访问服务	20	2
	安全与隐私	10	6

11.2.3 从应用开发环境与工具层分析我国的发展差距

在 e-Science 建设的 31 类组件中应用开发环境与工具层有 11 类组件, 我国的 e-Science 建设在这一层按项目数排名, 前 7 类组件分别是数据管理服务、工作流服务、可视化服务、模式与程序原型、数据挖掘与分析服务、网格管理服务和合作工具包 (图 11-3)。而从世界范围内来看, 应用开发环境与工具层涉及项目最多的前 7 类组件依次是数据管理服务、网格管理服务、数据挖掘与分析服务、可视化服务、合作工具包、虚拟与沉浸环境服务和工作流服务 (图 11-4)。

从图 11-3 和图 11-4 可以看出, 在这一层世界上投入项目最多的前 7 类组件中的 6 类也属于我国投入项目最多的 7 类组件之中, 这至少说明我国 e-Science 建设在应用开发环境与工具层上重点研究的组件基本上与世界范围内在这一层重点研究的组件相一致。但具体分析我国在这一层的前 7 类组件, 可以发现我国与世界上主要国家在这一层的建设上还存在一定的差距。我国在这一层建设上还主要研究基本的组件, 比如数据管理服务类组件、工作流服务类组件和可视化服务类组件, 是我国投入项目最多的三类组件, 但是这三类组件都属于这一层的基本功能。而我国对这一层提供高级功能的组件研究则相对较少, 比如, 网格管理服务类组件全世界总共有 71 个项目研究, 但我国只有织女星网格、中国气象应用网格和资源环境网格 3 个项目研究; 又如, 合作工具包类组件, 全世界共有 57 个项目研究, 而我国也只有 3 个项目研究; 此外, 数据挖掘与分析服务类组件全世界共有 68 个项目研究, 而我国只有 4 个项目研究。可见, 我国 e-Science 建设在应用开发环境与工具层提供高级功能的组件开发上投入还不够。

11.2.4 从具体应用层分析我国的发展差距

从我国的 30 个 e-Science 项目来看, 中国的 e-Science 具体应用领域主要包括生物、交通、教育、气象、航空制造、森林资源、新药研发和免疫学。其中, 除了生物学领域有 2 个项目外, 其他应用领域都只有 1 个项目。此外, 有些应用领域如农业、地球科学等在我国的 e-Science 应用领域中还依然是空白。尽管由于中国 e-Science 研究项目的总数与其他国家和地区之间存在较大的差距, 所以我们不能够仅仅从某学科或者专业领域相关项目的数量来评判中国 e-Science 在具体应用领域的发展差距, 但是我们却可以从现有项目的应用范围中看出我国在 e-Science 应用方面的发展差距。比如, 通过调查, 我们可以看到美国的 e-Science 项目已经广泛分布于天文、物理、生物、化学、工程和地球科学等多个方面, 而英国的 e-Science 研究则更是应用到了科学研究的各个角落。因此, 我们可以得出这样的结论: 中国的科学研究还没有完全接纳 e-Science 这一新的科研模式。换句话说, e-Science 在中国尚有广阔的应用空间有待进一步开拓。

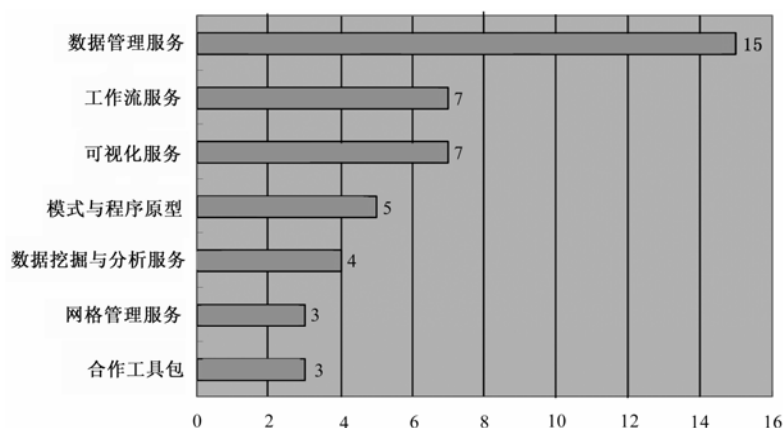


图 11-3 中国 e-Science 建设中应用开发环境与工具层各类组件涉及的项目数

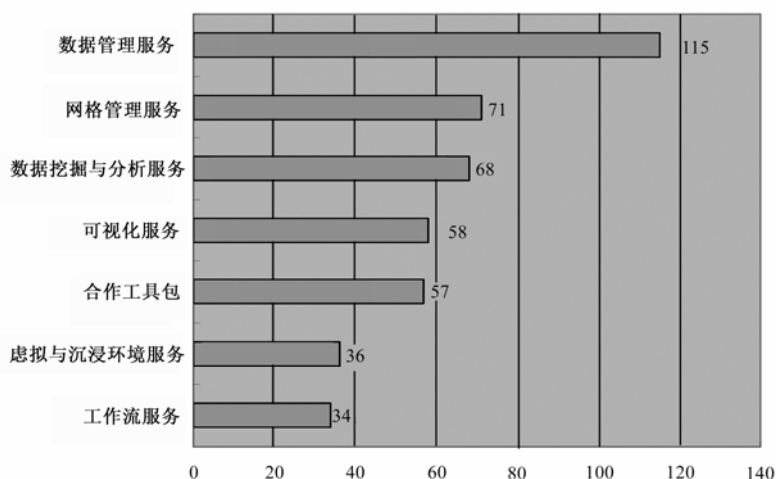


图 11-4 世界 e-Science 建设中应用开发环境与工具层各类组件涉及的项目数

11.3 中国发展e-Science策略分析

自英国于 2000 年首次提出 e-Science 概念并于 2001 年开始掀起 e-Science 建设热潮，继英国之后，欧盟、美国、日本等国家也先后投入大量资金、人力等资源到 e-Science 的研究与建设当中。迄今为止，国际 e-Science 发展已经表现出七大发展特征：e-Science 已经成为发达国家科研模式创新的方向；网格技术成为 e-Science 核心技术；基本形成统一的 e-Science 技术体系；大规模的合作成为各国 e-Science 建设的主要方式；各国 e-Science 建设与学科领域、应用紧密结合；各国 e-Science 规划实施各有特色；政府在 e-Science 建设中宏观规划和管理发挥主导作用。

这七大发展特征无疑可以为中国发展 e-Science 带来启示。

1. 进一步加强顶层设计、完善e-Science规划

我国在开展大规模 e-Science 建设的过程中,需要在《纲要》的基础上,进一步加强顶层设计、完善 e-Science 规划。以英国的统一规划、欧盟的试点平移拓展为参照。

在规划阶段,可以借鉴英国统一规划的模式,结合国家科技基础条件大平台的规划,由科技部牵头,制定中国的 e-Science 建设目标、建设任务以及实现途径等政策性指南,并启动中国 e-Science 建设计划,借鉴英国在开展 e-Science 建设的过程中实行核心计划与一般计划、核心计划实行阶段推进的规划模式。通过中国 e-Science 核心计划的实施,避免重复研究,协调建设,优势互补、规模研发,完成中国 e-Science 的基础设施建设。

在全面推广阶段,可以进一步借鉴欧盟的试点平移拓展模式。在核心计划充分实施的基础上,加强重点学科领域的 e-Science 建设,最后进行 e-Science 的正常运转、维护管理阶段,并将 e-Science 建设从东部沿海发达地区向西部落后地区全力推进。

2. 建立国家层面上的管理运行体系

借鉴 e-Science 的运行管理机制,在开展大规模 e-Science 建设的过程中,可以由科技部设立专款,设立由政府官员、网格科学家、业界代表者等组成的总顾问委员会,统一管理上、技术上指导全国的 e-Science 建设。

由于在网格建设阶段我们已经取了不错的成就,尤其是在网格中间件、应用开发环境与工具以及网格应用方面,我国已取得了不少成果,而这些成果恰恰又是发展 e-Science 所必备的基础设施,完全可以全部嵌入到我国大规模的 e-Science 建设中。因此,我国可以构建中国国家 e-Science 研究中心,以上海、北京、广东、浙江等在网格方面研究具有出色表现的地区为依托,建立多个 e-Science 地区研究中心,形成以国家 e-Science 研究中心为核心、以地区 e-Science 中心为支撑的 e-Science 完整体系。

为了保证 e-Science 的建设成效,需要仿照英国的做法,建立符合中国 e-Science 建设实情的评估体系与交流机制。一方面,从建设成果、人员培养等方面监督各个 e-Science 中心的运行绩效;另一方面,通过开展广泛的交流,将有助于国内 e-Science 建设经验的共享,进一步提高我国 e-Science 的建设效率。

此外,在合作机制方面,中国也应该加以充分重视。总之,通过完善的 e-Science 研究体系、科学合理的评估机制、保证共享的交流机制以及目标明确的合作机制,一个科学合理的 e-Science 管理运行体系将最终确立并发挥实效。

3. 进一步完善资源层建设

当前社会发展正经历着重大的转型,资源的占有和配置、创造和利用的优劣,不仅对国家科技竞争力有着关键的作用,而且也是 e-Science 建设的重要基础。

随着我国国家科技基础条件平台的建设,我国已经初步具备了 e-Science 建设五

层结构中的资源层雏形。资源层的建设主要包括以下五个方面：科学数据（指人类社会科技活动所产生的基本数据，以及按照不同需求而系统加工的数据产品和相关信息）、科技文献信息资源、自然资源（指经过长期自然形成的、对人类生存与可持续发展不可或缺的、为人类科技与生产活动提供基础物质，并对科技创新与经济发展起支撑作用的自然资源）、网络科技环境和科学仪器设备。随着我国 e-Science 建设马上开展，可以在已有的国家科技基础条件平台的基础上，进一步完善 e-Science 的资源层。

具体实施方案：通过国家层面的统一规划和部署实施，建成以国家为主导、行业主要部门负责参与的资源层。在国家科技基础条件平台的基础上，完善和维护已有的条件平台，如科学数据共享平台、科技文献信息资源与服务平台、自然资源平台和网络基础服务平台，同时构建和发展科学仪器设备平台。同时，在 e-Science 的发展过程中，不断动态更新和完善 e-Science 资源层。

4. 大力发展网格基础设施层

在网格基础设施层，我国应该继续扩大资源整合的范围，提高网络连接性能，并在 OGSA 框架的指导下，进行网格间的互连，真正实现 e-Science 整合全球资源的梦想。

另外，我国还应大力发展在现有网格基础设施基础上的具体应用网格的开发和建设，充分利用网格基础设施所提供的资源平台、服务平台，让资源真正能够为科学服务，为科研人员所用。

5. 重视网格中间件层的建设

在网格中间件层，我国虽然已经具有一定的研究能力，达到了先进的研究水平，并开发了数个拥有独立资助知识产权的网格操作系统。但在日后的研究发展过程中，仍然需要注意对国际标准的支持以及参与制定。因为只有这样，才能为我国网格技术完全与国际接轨，为全球网格系统整合，世界范围内 e-Science 的实现打下坚实的基础。另一方面，还应充分利用现有的开放源代码的研究成果，以避免重复开发。

而在具体网格中间件的组件研发过程中，应该在实现网格基本功能的基础上，对提供稳定、智能、灵活的网格服务这一目标进行努力，从而为科研人员提供更加强大的研究环境做出必要的准备。

6. 全面推进应用开发环境与工具层的建设

我国在未来 e-Science 的应用开发环境与工具层建设中应加强网格管理服务、数据挖掘与分析服务和合作工具包等组件类的研究开发力度。我国已有的 30 个 e-Science 项目已经对这一层的各类主要组件展开了研究，但是目前的研究的组件主要提供数据管理、可视化和 workflow 服务等基本的功能，而对网格管理服务、数据挖掘与分析服务和合作工具包等高级功能的组件研究则相对较少，而时序安排、监测

与诊断服务、仪器管理服务、虚拟与沉浸环境服务等组件我国则还没有项目对其进行研究。因此,我国在未来应加强对这一层提供高级功能的组件研究开发的力度。

7. 加强面向生命科学、医学、环境、农业等应用领域的e-Science建设

我国在未来应加强面向生命科学、医学、环境、农业科学等应用领域的 e-Science 建设。生命科学、环境和农业科学是我国“十一五”规划的重点学科领域,同时也是世界各国在 e-Science 建设中的主要应用领域,因此,我国的 e-Science 建设应加强这些领域的应用,以支撑这些学科领域的科学技术创新。同时,医学、工业技术等领域也是世界各国的建设重点,而我国目前还没有面向医学方面的 e-Science 研究,因此,在未来也应加强这些领域的建设,使 e-Science 建设真正做到以应用为导向。

8. 发展模式

(1) 坚持试点先行,横向拓展

我国的 e-Science 建设应坚持试点先行、横向拓展。借鉴英国和欧盟的经验,同时根据我国经济实力和科研实力的地区差异的国情,我国的 e-Science 建设应首先在经济发展较快和科研实力较强的环渤海湾地区、长三角地区和珠三角地区进行试点,在这三个地区的中心城市北京、上海和广州成立三个 e-Science 地区中心。在 e-Science 试点取得成功的基础上,将前期 e-Science 建设的成果横向拓展到我国的中部地区、西南地区、西北地区和东北地区,争取在 e-Science 建设的关键时期内在全国范围内建成十个 e-Science 地区中心,完成我国 e-Science 建设的国家级中心的布局。

(2) 层次推进的发展模式

我国在建设 e-Science 的过程中应坚持层次推进的发展模式。我国的 e-Science 建设应在国家 e-Science 顶层规划的指导下,以层次递进的方式来逐步而有序地推动整个进程。按照这种发展模式,我国应首先在国家级的层面上推动 e-Science 的建设,在国家级 e-Science 建设初步完成的基础上,进一步推进 e-Science 建设在省一级的布局,积极应用和推广国家级 e-Science 建设中的经验和成果,争取在我国主要省份完成 e-Science 建设的省一级布局,从而初步建成覆盖全国主要地区、层次分明的 e-Science 环境。

(3) 平衡推进,全面发展五层结构

我国的 e-Science 建设应遵循以应用为导向,平衡推进,全面发展 e-Science 五层结构的模式。五层结构是在全面研究世界各国 e-Science 建设后得出的,我国的 e-Science 建设应全面而均衡地发展五层结构中的各个层次,在各层的建设过程中可借鉴英国、美国和欧盟的经验,始终以应用为导向,使 e-Science 建设紧密与具体学科应用相结合,避免脱离实际应用需求的盲目推进,要在取得成绩的基础上平衡而稳步地推进我国的 e-Science 建设。

9. 建立以应用为导向的合作建设机制

(1) 组建联合建设团队

随着全球化的发展, e-Science 也在倡导交流与共享、合作与协作的精神, 同时通过本课题的项目调研数据说明, 大规模的合作已经成为当前 e-Science 建设的主要方式, 这种合作既包括国家之间的合作, 同时还包括的因此在我国 e-Science 建设中也应该积极倡导合作机制。

① 建立跨机构之间的合作建设机制

根据当前 e-Science 建设项目中, 所有承担机构的性质可以大致划分为 6 种类型: 政府部门、科研机构、高校、公司企业、协会组织、其他机构 133, 其中, 高校包括大学院系、大学研究中心和大学实验室; 其他机构包括图书馆、医疗机构和教育机构。

我国 e-Science 建设中, 主要参与机构是科研机构、大学和政府部门。科研机构在各层都发挥着积极的主导作用, 都占有绝对性优势, 其中在网格层作出的贡献最大; 大学在五层机构中也比较活跃, 涉猎各层, 但是作用要稍逊于科研机构; 中国政府也参与到具体应用层的建设中。

在国内 e-Science 的建设过程中, 社会参与合作力度仍然不够, 主要是以科研机构和大学为主体, 而公司企业和一些协会组织并没有发挥他们应有的一些贡献。虽然也有部分项目在尝试合作, 如中国教育科研网格与各大公司合作建设 China Grid 的公共平台, 并正在与包括英特尔、惠普、IBM 在内的公司展开关于平台方面的探讨和研究, 并准备开展合作, 但是这个公共平台目前在国际上并没有统一的标准, 也没有很好地完整实现。

因此, 在我国的 e-Science 建设中, 应该加强机构之间的合作, 建立以应用为导向的合作建设机制。政府部门在五层结构模型中, 发挥主导作用, 通过充分深入细致的研究、分析和遴选, 进行统筹规划, 鼓励国内各机构之间的合作, 尤其是应用机构与研发机构之间的合作, 组建以应用为导向的合作建设团队。通过合作, 使得资源共享、优势互补, 达到双赢, 从而产生整体效益和规模效益。在合作过程中, 要根据应用型机构的需求, 利用研发机构的研发和技术, 充分发挥各机构自身的优势, 有针对性地开展研究工作。

具体实施方案: 首先在国家层面对 e-Science 建设进行统一规划, 制定 e-Science 建设过程中的相关方案标准; 然后在国家的统一规划框架下, 以应用为导向, 根据应用和需求比较迫切的学科或领域建立 e-Science 建设项目, 最后根据具体的项目遴选合作建设团队, 包括应用型机构、研发机构、技术机构等, 该团队可以打破地域、学科、国家等限制, 根据项目的特点和需要选择合作机构, 同时可以在项目的建设过程中, 根据项目的建设进程和建设阶段重点, 进一步发展和完善合作团队, 动态组建该建设团队。

② 建立各层次均衡发展的合作建设机制

在 e-Science 的五层结构中,包括具体应用层、应用开发环境与工具层、网络中间件层、网络基础设施层和资源层。具体应用层直接面对用户,满足各种应用中对海量计算能力、远程数据获取和昂贵科学仪器设备的共享需求。具体应用层为用户提供一个应用服务的门户,通过该门户直接提交数据、访问远程资源和搜集研究结果;应用开发环境与工具层为 e-Science 开发人员和用户提供各种高水平的服务,该层主要为开发人员提供开发应用程序以及测试性能与可靠性的服务,为用户提供有效而方便的方式利用这些应用的服务;网络中间件层提供一般的网络应用服务和特定应用服务,这些核心服务包括远程处理管理、资源联合配置、存储访问、信息登记、安全、数据访问与转换、服务质量;网络基础设施层是支持 e-Science 各种应用的基础,主要是指高带宽、连通性好的通信网络,该层主要定义资源层各种数据交换的网络通信协议;资源层是整个 e-Science 框架的最底层,这一层由各种可获得的资源构成,包括超级计算机、计算机集群、存储网络、专业数据库服务、昂贵的科学仪器设备以及可视化设备。

通过对国际上 e-Science 和我国当前的 e-Science 建设的项目调研中发现,我国当前 e-Science 建设的发展主要是以网络中间件层为基础,以应用层为重点的建设方式。要最终实现 e-Science 共享与交流模式,就应该实现各层次均衡发展的合作建设机制。

在制定具体实施方案上,可由国家对 e-Science 建设进行统一规划,制定 e-Science 建设过程中的相关方案和标准的基础上,各层之间相互依赖,互为条件和补充,均衡发展五层结构中的各层;以具体应用层为切入点,根据国家层面对 e-Science 建设和规划,在具体的 e-Science 建设项目指导下来开展具体应用层的建设,进而以点带面,带动我国的整个 e-Science 建设;以应用开发环境与工具层为工具,根据具体应用层的需要,开发定制的开发环境和工具,为具体应用层提供服务;以网络中间件层为中介方法,根据具体应用层、应用开发环境与工具层和网络基础设施层的具体工作内容,提供五层结构中所需要的网络应用服务和特定应用服务;以网络基础设施层为技术基础,在我国网络环境建设的基础上,针对我国 e-Science 建设的总体方案和具体项目,根据国际上通行的标准,从国家层面统一标准,定义各种标准的网络通信协议,减少重复建设;以资源层为资源基础,包括科学数据、科技文献信息资源、自然资源、网络科技环境和科学仪器设备,为具体应用层、应用开发环境与工具层、网络中间件层和网络基础设施层提供资源。

(2) 开展国际合作

大规模的合作已经成为当前 e-Science 建设的主要方式,我国作为科技大国在发展 e-Science 建设也应该打破国家的界限,走出中国,与世界范围内的各国进行合作。我国当前的 e-Science 建设中已经与世界上的有些国家建立了良好的合作关系,但是与全球合作的大趋势相比,仍然需要进一步加强和努力。

具体实施方案:主要从三个层面来进行,首先是国家层面,可以由国家科技发

展的主要职能部门，如科技部，对外与 e-Science 建设的主要国家负责部门建立良好长期的合作关系，包括资源共享、人员交流与培养、召开国际会议、共同承担项目等方式，对内组织各方力量，即时向国际上公布我国 e-Science 建设进展，在确保国家机密的同时，扩大我国在国际上的知名度和影响力；其次是机构层面，国内各机构在国家层面的领导下，可以与国际上相似的机构之间达成良好的 e-Science 建设长期合作机制，包括技术交流、人员交流与培养，共同承担项目等；最后是项目层面，根据具体的 e-Science 建设项目，与国际上在该领域上比较成熟的项目组和机构之间展开合作，包括经验交流、人员交流或者共同承担项目等。

10. 加强中国e-Science人才队伍建设

在网络建设的过程中，中国已经积累了一定的 e-Science 人才建设经验、并形成了初步的 e-Science 人才培养机制。如采用“走出去、请进来”的培训模式，一方面，邀请国外顶尖网格专家、e-Science 研究者到国内相关项目担任技术顾问、参会发言；另一方面，通过 e-Science 国际会议、网格会议或者是项目合作的方式，让有关人员有机会到国外亲身到体验 e-Science 的建设模式。

尽管如此，国内在 e-Science 人才建设方面目前尚未形成规模，而且高端人才培养不多。因此，随着我国 e-Science 建设的深入，e-Science 人才队伍的建设已经是一个必然的趋势。在此过程中，必须坚持两大建设原则。

(1) 加强高端人才的培养

e-Science 的顺利实施，涉及多项先进技术，如网格技术，技术复杂度、综合度都很高，缺乏高端人才的培养，都将阻碍 e-Science 进步。在此过程中，为了保证我国 e-Science 研究与建设与世界同步，需要加强高端人才的培养。

(2) 以项目带动、注重实践能力

在培养 e-Science 人才的过程中，需要从动手能力方面进行加强。这主要可以通过让培训人员实质参与项目或者是让培训人员到国外去访问学习 e-Science 相关建设经验的方式来实现。例如，可以让一些专门从事网格技术研究或者 e-Science 专门研究的机构或实验室提供博士后学习机会，也可以让受训人员到英国国家 e-Science 研究中心去做访问学者等。

第四篇 支撑服务篇

科学研究以对信息资源的全面占有和准确分析为前提。长期以来，文献情报机构已经成为科研活动不可或缺的支撑和保障。近年来，新技术的不断出现和迅速普及导致信息资源数量的急剧膨胀和质量的良莠不齐，信息泛滥、信息污染等使得科学研究对文献情报机构的依赖性日益增强。在 e-Science 环境下，文献情报机构的服务保障功能不仅依旧存在，而且还会继续加强。应该看到，e-Science 目前的研究重点仍是网格基础平台的建设及其在科学研究中的应用以及对数据资源的共享等方面，而且目前的研究基本集中在技术层面上，参与者多是计算机专家和科研人员，文献情报机构尚未介入其中。

e-Science 环境下的科研活动要求文献情报机构继续提供持续的、强大的支持。e-Science 将使科学研究超越对传统文献情报服务的依赖，从而给文献情报机构的发展带来一系列新的挑战。

尽管 e-Science 正处在孕育和形成阶段，但新的需求环境正日渐形成，为了跟上服务环境的变化，使文献情报工作及时适应新环境，有必要从新环境初露端倪时就积极地寻找对策，寻找文献情报机构融入新的服务环境的切入点和生长点。因此，本篇拟就 e-Science 环境下的文献情报机构发展战略进行探讨，期望能够揭示这种新服务环境下文献情报机构的发展方向，为文献情报工作制定战略决策提供参考和借鉴。

第 12 章 e-Science环境下文献情报机构发展分析

12.1 e-Science对文献情报机构工作环境的影响

12.1.1 e-Science环境下科学研究的过程分析

e-Science 对文献情报机构的影响是通过它对科学研究的影响间接引发的，通过分析科学研究的过程，一方面，可以更清楚地认识到在 e-Science 环境下文献情报机构需要并能够在科学研究的哪些环节，以何种形式更有效的融入科研过程，为之服务。另一方面，只有了解和掌握科学研究和科学发现的过程，才能在服务中有的放矢，向科研用户提供最有针对性的、用户最需要的产品和服务，从而有效地为科学研究服务。e-Science 环境下科学研究的过程如图 12-1 所示。

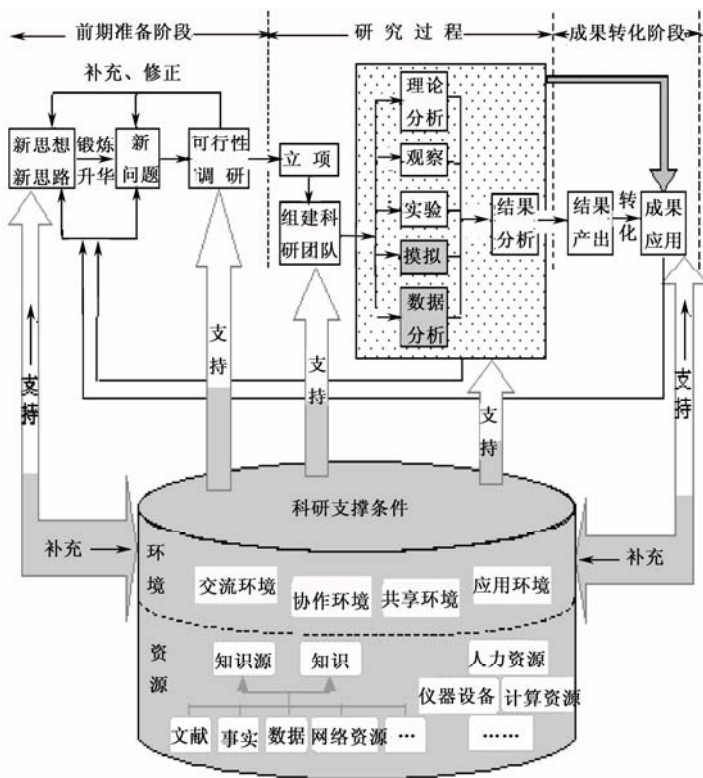


图 12-1 e-Science 环境下科学研究的过程

科学研究的过程可简单的划分为三个阶段：

- 准备阶段，即确定一项科学研究是否有研究的价值和必要性。
- 研究阶段，在确定开展一项科学研究之后，组建科研团队，运用各种手段实现既定目标，如形成新理论、新学说，发现新物质、新规律，找到新方法，开发新工具。
- 转化阶段，将科研成果应用于现实中，这是科研成果价值实现的途径。

当前，文献情报机构的服务活动基本处于支撑环境中，围绕着以文献型资源为主的信息资源的管理和利用展开的，它还处在科学研究的外围，虽然也会在某些时候、某些环节介入科研活动，但并非融入其中，文献情报服务与科学研究基本上是分离的。在 e-Science 环境下，科学研究的各个阶段都需要各种信息、仪器设备等资源的支持，而且这种支撑的力度和紧密性将比传统环境下更强。因此，科学研究需要文献情报服务在各个阶段融入科研活动之中，这种融入将是人员、产品、平台、服务的全方位融入。

12.1.2 e-Science给科学研究带来的新变化

e-Science 环境为科研活动提供了更先进、更完善的研究工具和平台，它给科学研究带来的影响主要体现在如下几个方面：

（1）使科学研究的方法发生了革命性的改变

理论和实验是支持传统科学研究的两大基本方法，科学家通过思考和实践，不断地认识世界、掌握客观规律，并应用所掌握的知识来指导实践、创造发明。e-Science 的出现为科研活动带来了方法上的革命，它极大的改善了传统的科研方法，如计算和测度精度的提高，同时带来了新的方法——模拟和仿真，使人类的触角可以伸向在传统的实践方法下无法触及的领域，如超高温环境下的实验，从而使得科学研究的深度和广度、质量和效率都有了新的飞跃。

（2）使科学研究成为一种开放性的、全球合作的、协同式的研究

e-Science 意味着通过全球分布协作，使用海量数据集合、万亿次级的计算资源和高性能的可视化提高科学研究效率，它使得科学研究的大规模协同和合作可以容易地、长久地持续下去，成为科学研究的一种基本方式。e-Science 环境下，虚拟组织成为科学研究的基本组织形态，不同地域、不同机构、不同领域的科研人员将因为共同的问题虚拟地聚集起来，利用分布的资源进行科学研究。

（3）实现了资源的全面共享

e-Science 的出现将消除资源孤岛，它将通过网格中间件和高速网络实现各种异构资源的虚拟集成，使网络成为一台超级计算机，从而实现计算资源、仪器设备、数据资源、信息资源、知识资源、人力资源等资源的全面共享。例如，科学家可以存取存储在不同地理位置的超大规模的数据集，对它们进行分析和实验。

(4) 科学研究的起点不再一定是观察和实验

先进技术使得详细保存前人的研究过程和研究成果成为可能,后来的研究不仅仅建立在前人的研究成果之上,还可以根据自己的需要从前人的研究过程中的任何一点开始,沿着不同的方向继续研究,结果可能会有不同的发现。正如 DoubleTwist 生物公司所做的那样,他们摒弃了进行 DNA 分子和生物研究必须的瓶瓶罐罐,其分析工作全部基于合作伙伴提供的基因序列,在此基础上构建数据库,并用自己开发的分析系统对基因进行分析。

(5) 动摇了传统的科学交流的模式

e-Science 的出现将会对传统的以正式出版为主导的学术交流体系造成冲击。在 e-Science 环境下,学术交流的手段多样化、途径便捷化,科学交流空间数字化、网络化、虚拟化,科学交流的模式将发生重构,数字化网络化的学术交流体系将最终确立。

12.2 e-Science环境下文献情报机构的服务对象分析

12.2.1 用户类型分析

由于不同类型的用户有不同的需求,因此,对用户类型的细分有助于把握用户需求的重点和要点,以便更有针对性地提供服务。e-Science 环境下文献情报机构的服务对象包括所有科研人员和与科研活动相关的一切辅助人员,根据在科研活动中的作用不同对这一用户群体进行细分,可以获得以下几种类型的用户。

(1) 科研决策者

科研决策者是引领科学研究发展方向的主导力量,其需求体现在两个方面:一方面是前瞻性、战略性情报,如前瞻性的学科战略情报研究;另一方面是新颖、全面的原始资料。在很多时候,文献情报机构提供的综述从深度、综合性、创新性等方面无法达到决策者的要求,而且很可能由于对专业知识的欠缺和对最新发展动态的把握不足而过滤掉了有价值的信息,因此,决策者和他们的顾问团需要更多的原始资料做分析,支持自己的决策。

(2) 科研管理者

科研管理者是科研活动中各个部门、各个环节之间联系的纽带,他们组织、协调、管理着科研活动,维系着科研活动的正常运转和科研秩序的稳定。他们的需求并不一定很专深,但涉及的范围无论是学科跨度还是历史跨度都可能很广泛。

(3) 科研实践者

科研实践者是指从事具体科研工作的科研人员,这部分用户的需求大多集中在具体的、细节性的资料上和对具体问题的解答上,目的性很强。

(4) 科研顾问

科研顾问是科研决策者的参谋，帮助决策者把握正确的发展方向，他们与决策者的需求是一致的，这里不再赘述。

(5) 科研活动中涉及的其他人员

这部分人员对科研活动起着支撑和辅助的作用，如法律顾问（在科研项目的委托和立项过程中、科研成果转化过程中、专利成果的申请和维持过程中都需要法律顾问保证过程和结果的合法，保护自身的权益和利益）、技术平台的提供者和维护者等。这部分用户的需求是零散的，涉及的领域广，有时学科跨度会很大。

无论何种类型的用户，尽管他们的工作内容不同，但他们都是组成 e-Science 环境下科研活动不可缺少的部分。用户类型的多样性决定了其需求的多样性，也决定了文献情报服务系统的灵活性，文献情报服务需要考虑所有类型用户的需求，并针对不同用户的需求提供有针对性的服务。

12.2.2 用户信息需求的变化

1. 用户信息行为的变化

e-Science 的出现给科学研究带来的变革引导着科研用户改变其创造知识的行为，从而导致其信息行为的变化。信息获取的时空障碍将被全面突破，数字化获取正在成为用户获取信息的主渠道，正如美国国家科学基金会在其报告《利用网络基础设施变革科学与工程研究》中指出的那样：在越来越多的研究领域中，科研人员对最新科学发现和成果方面信息的获取途径首先是万维网，然后是预印本和学术会议，最后才是科技期刊。同时，自动化操作将成为用户利用信息的主要方式，用户将更多的依赖网格远程获取和共享信息，依赖计算机软件和共享的计算资源、仪器设备组织、管理、分析、利用数据和信息。

2. 用户需求内容的多样化

e-Science 环境下，新技术的应用将使资源体系的构成情况发生变化。一方面各种资源的数量比例将被改写，如大批高性能仪器设备的联机使用和协同工作将引发数据资源数量的急剧膨胀，使数据资源在资源体系中所占比例大幅度提高。另一方面，新型资源不断产生并迅速积累，如超媒体的出现，同时更多的资源类型（如计算资源）有可能纳入文献情报服务的掌控范围，使资源的类型更加丰富。资源类型的多样化和构成比例的变化将引发它们在科学研究中的作用和价值的变化，科学研究所依赖的将不再仅仅是文献型资源，而向着多类型方向发展，这将引发用户需求内容上的多样化，用户需求将跳出传统的以文献型为主的模式，向着多类型方向发展。

3. 用户需求内容的层次化和深化

e-Science 环境下的海量信息和资源多类型化使用户的需求无法再停留在原始资源的获取上,也不再满足于对资源中所蕴涵的信息的获取,而是不断深化,直指文献、数据、信息等资源中蕴涵的知识、概念、知识元及相互之间的关系等深层次的对象,形成多层次的信息需求。

4. 用户需求的重复性和延续性

e-Science 环境为科学研究创造的便利将使科学研究从宏观和微观向着更广、更深的方向发展,科学研究日益复杂化。这种向复杂性方向的演化使得一些科学研究被细化,分割成前后相继或并行发展的若干项目,而这些研究之间存在着明显的相似性和相关性,因此,其用户需求具有很强的重复性(研究之间并行)和延续性(研究之间前后相继),研究的过程、方法、工具可以互相参考、借鉴,甚至互用。

5. 用户享用服务的虚拟化

e-Science 环境下,用户通过虚拟的网络实现彼此之间的交流和各类资源的共享,因此对服务的享用也向着虚拟化方向发展,希望能够通过网络虚拟的享受服务,希望服务的提供智能化、简单化、个性化、专业化、知识化。

12.3 e-Science环境下的文献情报服务

12.3.1 e-Science环境下的文献情报服务及其定位

1. e-Science环境下的文献情报服务范畴

传统的文献情报服务习惯于按系统、地域或机构划分服务对象和范围,每一个文献情报机构都具有相对明确的服务对象,如某个地区、系统或单位。

e-Science 环境下,由于科学研究的跨学科、跨机构、跨地区,将使得某个具体的文献情报机构的服务对象和服务范围不断外化,其服务对象将可能是位于不同地域或机构,从事不同学科研究的科研人员。因此,以往按学科或机构组织资源和服务的做法将不再适用于 e-Science 环境下的文献情报服务。

2. e-Science环境下的文献情报服务手段

传统环境下,文献情报服务机构各自独立的完成为科研提供服务的任务,目前已有的合作基本停留在资源的共建共享方面,还没有深入到服务的其他方面。网络的功能主要是检索资源和传递结果,还没有真正实现实时交互式的服务。

e-Science 环境下, 协作式服务将成为服务的主要方式, 不仅仅是资源建设上的协作, 还包括人员、设备、服务方面的协作和共享。文献情报服务的一些传统服务手段将得到发展, 如个性化服务将为用户建立一个包含用户所需的任意资源和服务的动态的、可随时更新、随时存取的, 与用户个人的知识管理系统集成的图书馆。同时, 一些新的服务手段也将出现, 如对科研过程、交流过程进行记录、检索以及对任意片段随机存取和播放。

3. e-Science环境下文献情报服务的处理对象

传统环境下, 文献情报服务依靠和管理的主要对象是文献型资源, 尽管近年来事实型和数值型资源也在文献情报服务的工作过程中出现, 但它们远没有占居主导地位。

e-Science 的出现会导致资源类型的多样化, 除了传统的文献型资源外, 各种新型资源将出现并逐渐成为文献情报服务所依靠和管理的主要对象。e-Science 的目标之一是实现计算资源、数据资源、信息资源、知识资源、人力资源等资源的全面共享, 这些资源都将成为 e-Science 环境下文献情报服务所依靠和管理的对象, 如科学数据、化学试剂、各种仪器设备。不仅如此, 从信息资源的载体形式上来说, e-Science 环境下原生数字化资源将急剧增多, 成为资源存在和存储的主要形式。

4. e-Science环境下文献情报服务的运行方式

传统环境下, 文献情报工作发挥的是一种中介作用, 它常常游离在科学研究之外为科研提供服务, 落入“召之即来、挥之即去”的尴尬处境, 处于第三方的地位。

e-Science 环境下, 科学研究要求的是一种嵌入式的服务, 是融入科学研究之中、融入知识创造过程之中的服务。文献情报工作将在科学研究的许多环节上承担原先由科研人员完成的工作, 文献情报工作者将成为科研人员的一部分, 与科研人员形成一种合作者 (partner) 的关系, 而不再是科研信息的中介。

总之, e-Science 环境下, 文献情报服务将成为“科学研究核心机制的一部分”, 它应该是科研活动自然而然的组成部分, 只要科研活动在进行, 就有文献情报工作渗入其中, 二者之间的交融使得利用现行的标准已经不能够在 e-Science 环境下将文献情报工作从科研活动中分离出来。

e-Science 环境下文献情报工作者将以科研人员合作者的身份出现, 而不再是依附者。科研人员的工作中将掺杂着文献情报工作, 如科研人员在科研过程中按照文献情报服务所规范的要求、利用独特的工具记录实验的过程, 以实现在未来某个需要时刻共享、检索和重用这个过程。文献情报工作者的工作中也存在着科研工作的成分, 如帮助科研人员进行实验的准备。

12.3.2 e-Science环境下文献情报服务的作用

e-Science 环境下，文献情报服务不仅延续和发展传统文献情报服务的作用，还将解决 e-Science 建设和发展过程中遇到的一些新问题。

1. e-Science资源的管理

e-Science 的核心是资源的共享，它的各种目标也是通过各类资源的共享实现的。这里所说的资源并不限于文献情报服务传统意义上的信息资源，而具有更广的涵盖范围，e-Science 环境下可共享的资源可以归纳为如下几类：

- 计算资源：计算力，是计算设备所提供的计算能力；
- 仪器设备：包括各种实验仪器、存储设备；
- 通信资源：支持 e-Science 环境下不同节点之间彼此通信的相关设备，如电缆、交换设备、联网设备等；
- 信息资源：包括传统的文献、事实等资源；
- 数据资源：各种实验、模拟、仿真、分析产生的海量数据；
- 软件资源：用于处理各种资源、实现各种功能的软件工具；
- 人力资源：参与科学研究的一切人员；
- 知识资源：从各种其他资源（包括从人脑）中抽取的知识；
- 服务资源：可用的各种服务；
- 其他资源：科学研究中用到的，不包含在上述几类中的资源。

e-Science 环境下“资源共享”中的“资源”指的是科学研究需要用到的一切资源，对这些资源的共享会涉及方方面面的问题，技术建立的仅仅是一个共享的平台，而对其他问题的解决需要技术以外的力量。共享的前提是有效的管理资源，包括全面的收集、序化、整合和适当的处理，而不同类型的资源由于其属性、特点的不同，这些过程会不尽相同，遇到的问题也会不同。以往的文献情报服务工作主要是围绕信息资源展开的，而在 e-Science 环境中，文献情报工作将介入上述各种资源的共享活动之中，在这些资源的管理与共享过程中发挥作用。

2. 解决e-Science建设和发展过程中遇到的一些问题

e-Science 在建立过程中会遇到一系列问题，中国国家网络总体专家组组长钱德沛教授将网络面临的问题归纳为 9 个方面——计算模型、存储模型、信息模型、资源管理、服务、安全、普适计算、标准、运营。要解决这些问题，仅仅依靠技术力量是不够的，如信息模型、资源的管理、标准的建立涉及资源的组织、表示、编码、信息的交换、动态调用等方面的问题，这些问题的解决是文献情报工作者的责任。

3. 知识管理

e-Science 环境下, 知识将成为科研用户的主要需求, 知识服务将成为文献情报服务的核心服务, 因此文献情报工作者将实现从资源管理者、信息提供者向知识管理者的角色转变, 承担起知识管理的责任。

知识管理包括对知识的获取、存储、转化、共享、利用、长期保存, 各种知识管理工具的开发, 若干异构或同构的知识管理系统的集成等方面。知识的获取需要大量的信息等资源的支撑, 而对获取到的知识的管理、软件工具的开发、集成则需要相关的理论、方法和技术的指导。在知识管理的过程中, 文献情报服务将与其他的知识管理者合作, 通过它在信息管理中积累的经验、技术、资源为知识管理提供支持, 将个体的知识管理系统、组织的知识管理系统、社会的知识管理系统有机集成, 并以此为基础为用户提供知识服务。文献情报服务将融入知识创造的各个环节, 从宏观和微观两个层面上发挥作用, 成为知识转化与再生、创造中的催化剂和润滑剂。

4. 科研前导

科研前导是传统文献情报服务作用的延伸。在 e-Science 环境下, 一方面, 各种资源的大量涌现和科研节奏的加快使得科研人员在捕捉学科发展方向时变得日益困难, 难以抉择, 科研人员需要文献情报服务帮助把握科研发展的方向。另一方面, e-Science 环境为人们进行科研发展前景的研究和预测提供了更多、更有效的信息、数据和更强大的工具, 这使得文献情报服务可以更好地介入这项工作, 成为科学发展的前导员和领航员, 引领科学发展的方向。

5. 科研助理

e-Science 环境下, 科研过程中非核心环节的工作量骤增, 各种资源的生产速度和更新速度将越来越快, 知识的老化速度将不断加快, 科研人员已无力承担如此庞大的资源体系的组织和过滤工作, 需要文献情报工作者发挥科研助理的作用, 代替科研人员完成一些外围的、非核心的科研工作, 帮助其完成科研过程的部分环节。

6. 媒介作用

审视科学技术的发展历史可以看出, 知识创新有赖于对现有知识的充分吸收和利用, 这是科学技术发展的连续性和继承性所决定的。而这种吸收和利用的前提条件之一是对知识的载体——各种资源的有效管理和传播。

在 e-Science 环境下, 资源类型和保存格式呈现多样化, 资源的生产更加随意, 资源的数量日益庞大, 这些都使得对资源的管理和传播更加困难。因此, 文献情报工作应该承担起资源及其内含知识的传播媒介作用, 成为信息、知识传播的桥梁和媒介。

12.3.3 e-Science环境下文献情报服务工作的指导原则

指导原则是指引文献情报服务工作的坐标，它从宏观上规定了文献情报服务工作的方向，规范了开展文献情报服务工作的主线。指导原则并不是一成不变的，它会随着内外部环境的变化而动态调整。“在发展中调整，在调整中发展”是对待指导原则应持有的态度。

1. 以用户为中心原则

在图书馆的办馆理念从“书本位”向“人本位”转变之后，“用户需求驱动”便成为长久以来指导文献情报工作的首要原则，在 e-Science 环境下，这一原则仍将是指导文献情报工作的基本原则。

在 e-Science 环境下，目标用户群的类型、范围将会扩展，用户需求呈现多样化、复杂化、动态性、潜在性，因此用户需求的满足将更加困难。文献情报工作不仅要满足用户现实的需求，更要激发潜在的需求，要能够在稳定用户的同时“创造”用户，进一步发展用户群，扩大文献情报服务的生存地带和发展环境，实现文献情报服务的可持续发展。

2. 隐性化原则

e-Science 环境下的文献情报服务将完全融入科研过程之中，成为科研活动的一部分。从用户角度来说，已经没有什么“文献情报工作”、“文献情报工作者”的概念了，文献情报工作完全处于后台，隐于科研活动之中。

3. 协同性原则

e-Science 环境下的科学研究将是由全球科研人员协同完成的，因此，为其服务的文献情报工作要适应这种环境，也应该是协同性的。这种协同不仅仅是文献情报机构之间的协作，还有文献情报机构与科研用户、科研团队之间的协作，与其他科研支撑系统之间的协作，以及与其他类型的服务提供者的协作。文献情报工作者既要树立协同工作的思想和理念，又要培养协同工作的能力，以保证协同工作的高效开展。

4. 灵活性原则

e-Science 创建了高效的科学研究环境，在这种环境中，科学研究的节奏将会加快，知识创造的周期将会缩短，而学科之间的交叉融合也会随之加剧，这些都会使新问题的出现具有更强的不确定性——时间难以预测（不知道会出现在科学研究的哪个阶段），背景难以预测（不知道会出现在哪个学科、哪个领域的内部、边缘或交叉地带），性质难以预测（是否是文献情报服务要解决的问题已经没有明确的界定，许多问题交织在一起，需要文献情报服务与其他服务协同工作来解决）。

面对这种多变的服务环境，文献情报工作（者）要具有灵活的应对能力和策略。

要能够根据用户需求的动态变化灵活的驾驭和整合文献情报工作，实现资源的动态组合、人员的动态组合、工具的动态组合、服务的动态组合，以变应变，满足用户不断发展变化的需求。

5. 阵地化原则

在 e-Science 环境下，用户需求将更集中的体现在科学研究的过程之中，体现在知识创造的过程之中。文献情报工作要面向任务、面向课题，建立服务阵地，以提高服务的针对性。面向课题和任务意味着资源的组织、人员的配置、服务内容的策划、服务工具的选择都将面向课题和任务，其实质是面向用户需求，面向用户问题的解决。

6. 整体性原则

e-Science 环境下的文献情报服务以虚拟组织为实现的主体，以任务和课题作为服务的前沿开展服务，然后以学科和机构为汇聚中心，将为任务、课题提供的服务结果按学科、机构逐层汇聚，最终形成一个有机联系的整体。这个整体使得面向任务的不同服务之间建立起联系，使得彼此之间的服务过程和服务结果能够借鉴、参考、共享和重用。

7. 技术上的先进性与实用性相结合原则

随着技术的发展，人们越来越感到，必须建立以技术为支撑主体的知识服务机构来适应 e-Science 环境。e-Science 是一个高度技术依赖型的环境，这就决定了文献情报服务要在这个环境中生存，也必须高度依赖各种先进技术。因此，文献情报服务的实现，技术基础是必需的、决定性的，需要充分利用各种先进的技术搭建服务的平台、实现服务的传递。

文献情报服务在引进新技术的时候，要遵从先进性与实用性相结合的原则。在理论上要具有前瞻性，要认真研究新出现的先进技术和仍不成熟的先进技术的实用性，在技术发展的初期就了解它、掌握它，引导它的发展，实现为我所用。但是在实践中要以实用性为前提，在选用先进技术的时候要注意选用那些具有一定成熟度和实用价值的技术，而不要不切实际地一味追求技术的新颖和先进。

12.4 e-Science环境对数字图书馆的影响

12.4.1 e-Science对数字图书馆的积极影响——以SRB为例

1. SRB概述

SRB(Storage Resource Broker)是美国加州大学圣地亚哥分校的 SDSC(San Diego Super Computer Center)开发的一个面向网格环境下异构数据资源处理的数据网格软件。它为不同类型的存储设备提供统一的存取界面，屏蔽存储系统的异构特性，支

持广域网络环境下多种数据源的访问，为数字图书馆、持久保存系统和资源集管理系统提供了方便的数据存取功能。

由于 SRB 是一个联合的服务器系统，每一个 SRB 服务器管理一系列的存储资源，因此，它可连接各种分布式异构的资源，如来自 HPSS（High Performance Storage System）、UniTree 和 ADSM（Adstar Distributed Storage Management）等系统中的资源。SRB 将存储在多种异构系统中的信息资源组织为逻辑资源集合，为用户对数据资源的存取提供了便利的方式。现今的数据集和资源的存取都以数据集和资源的名称和物理地点为基础，SRB 与 MCAT（Metadata Catalogue）联合，改变了这一现状，实现了基于属性的存取，在获取数据集和资源时，用户根本无需知道数据集和资源的位置。

2. SRB技术特征

- 定位透明：用户能链接到任何一个 SRB 服务器来存取数据，可以通过任何一个全球唯一标识或者资源集合属性发现数据集。
- 可靠性和可用性：在不同 SRB 服务器提供的负载平衡控制下，数据可能在不同主机的不同存储系统中被复制。不同的存储系统可能会在不同的安全协议之下运行在不同的主机上，通过使用某个单一的登录环境和存取控制列表维持每一个数字化实体。
- 容错能力：当原始的存储系统不可获得时，系统自动重新定位其他单个存储系统中原始资料的某个复制品，并通过全球唯一标识进行数据存取。
- 持续性：通过递归的目录运转命令，数据能够被复制到新的存储系统中，而且不需要改变数据名称。这使数据集合转移到新的资源时不会影响到数据存取。

3. SRB在数字图书馆中的应用

对于分布式数据管理而言，SRB 是一个可靠的通用系统，这使得它能够在网格环境中获得广阔的应用前景。目前，SRB 系统已经被五十多个单位使用。它为创建跨越多管理领域、多类型存储系统、多类型数据存取环境的数据管理系统创造了核心基础结构。SRB 在数字图书馆的应用潜力表现在：SRB 技术可用于处理数字图书馆的无缝认证问题。数字图书馆使用 ID 管理数字化实体，从远程存储体中存取数据时，SRB 开发的相关技术能够管理用户对数据处理环境的认证、用户对存取某个数字化实体的认证以及远程记录的数据处理系统的认证。

作为数据网格的代表，SRB 系统为数字图书馆的资源组织与管理提供了良好的解决方案。目前利用 SRB 系统建设的数字图书馆主要项目有 SDLP（Simple Digital Library Interoperability Protocol）、NSDL（The National Science Digital Library）、NARA（the National Archives and Records Administration）、SIO Digital Libraries 等。

总之，SRB 技术成果在数字图书馆的成功应用仅是 e-Science 积极影响数字图书馆的一个例证，随着建设的深入，e-Science 将会进一步提升数字图书馆的资源处理

能力并最终提高其服务能力。

12.4.2 积极应对e-Science的数字图书馆

近年来，国际上已经开始对以网格技术为基础的数字图书馆技术进行探索，提出了数字图书馆建设的新想法与新思路。

在欧洲，由欧盟资助 630 万欧元、为期三年的基于网格技术的数字图书馆原型项目——DILIGENT (A Digital Library Infrastructure on Grid Enabled Technology) 也于 2004 年 9 月正式启动。该项目旨在“创建一个先进的试验床，使 e-Science 环境下动态虚拟组织的成员能够存取共享的知识，并以安全、协调、动态和低成本、高效益的方式进行合作。”

在中国，清华同方公司研究推出了 CNKI 网格资源共享平台，对网格技术在数字图书馆中的应用做出了大胆的尝试。该建设项目的建设者提出了知识网格操作系统的实现方法、网格资源共享平台的设计框架，并在此基础上提出了图书情报界依托网格资源共享平台开展知识服务的模式。CALIS 正在逐步推动数字图书馆软件系统向网格方向发展，在新一代数字图书馆网格发展目标的基础上，研究自身独特的数字图书馆网格产品。另外，目前中国国家图书馆正在积极展开文化信息网格的“缩比实验”，也即是，将文化信息网格总体模型按比例缩小并构建出在结构和功能上都比较完整的缩比模型。该项实验旨在研究文化信息网格的数学模型和系统模型，进而探索解决文化信息网格的核心技术问题，并建立文化信息网格的原型系统，为国家数字图书馆工程的顺利实施奠定基础。面对 e-Science 的到来，CSDL (Chinese National Science Digital Library) 也因时而变，将自身的服务定位调整为，致力于全球科研信息的集成和互联，将持续、可靠地支持全院的数字化科研环境。

面对 e-Science 的到来，数字图书馆正采取措施积极应对。在 e-Science 环境下，数字图书馆应该以何种方式提供何种贴切的、最能满足 e-Science 需求的服务，这需要深入研究 e-Science 环境下的用户需求，需要明了需求变化，更需要考虑 e-Science 给数字图书馆的资源、技术、服务、人员及政策标准等方面带来的影响。

12.4.3 服务e-Science的数字图书馆

Liz Lyon 于 2003 年指出，当前，英国数字图书馆界与 e-Science 研究机构之间存在四种合作动向：UKOLN (UK Office for Library Networking) 和英国国家 e-Science 中心联合召开专题研讨会；成立数据管理中心；启动 eBank 项目；成立国家文本挖掘中心。

1. eBank项目

eBank 由 JISC (the Joint Information Systems Committee) 和 EPSRC (the

Engineering and Physical Sciences Research Council) 共同资助。该项目的目标是, 建立 e-Research 数据、学术交流以及学习资源之间的联系。它通过使用诸如 OAI-PMH 这样的一些通用技术来探索、研究如何将科学研究的数据集无缝集成到数字图书馆中。图 12-2 展示了数字图书馆在 eBank 项目中的重要位置。

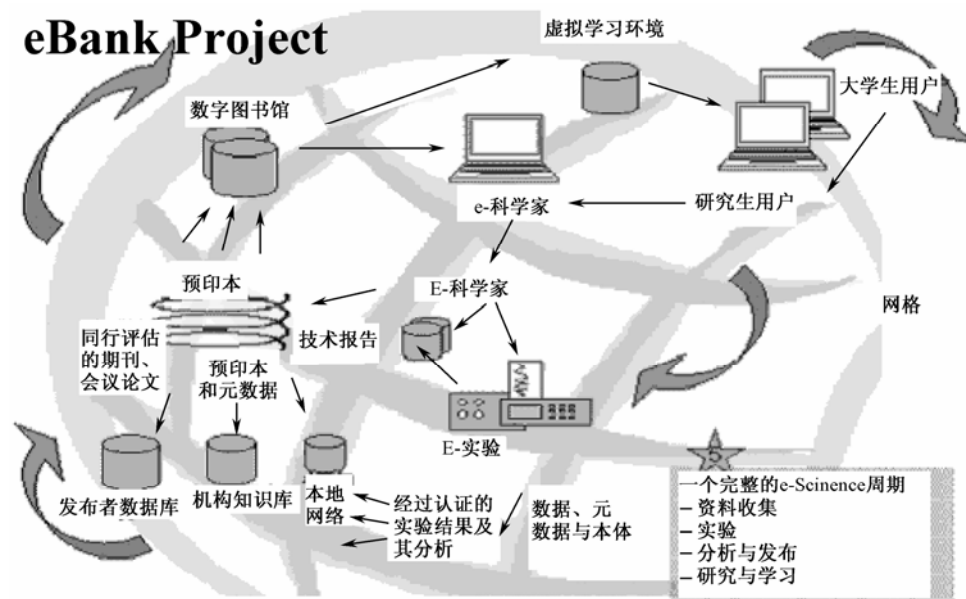


图 12-2 eBank 的体系模型

事实上, eBank 的建设也证明了 e-Science 环境下数字图书馆元数据技术与理念极其重要。除了网格技术, 语义网 (semantic web) 技术也是发展 e-Science 的中坚支撑技术, 而该技术的直接基础就是元数据技术。学者 Henry S.Thompson 在 2002 年全球网格论坛上指出, 语义网必须以元数据为基础。借助于元数据, 资源服务商才有可能对一个文档或者一项资源的内容或者是使用对象进行描述, 并可以让搜索引擎在语义信息的协助下更好的运作。

2. DCC项目

在未来五年内, e-Science 项目产生的科学数据将超过人类史上已有的数据总量。出于为当前与未来数据提供维护与增加价值的需要, 英国 e-Science 核心计划与 JISC 开展合作, 就数据掌管、数据保存和开放获取等重要问题共同进行探讨与研究, 并在此基础上建立了 DCC (Data Curation Center)。在 e-Science 的运行过程中, 数字图书馆将充当一个数据库的角色, 专门用于对各种网络内容、元数据、衍生数据进行存储, 并随时提供给数据研究人员使用。

第 13 章 e-Science环境下文献情报机构的服务模式

13.1 e-Science环境下文献情报机构的服务模式

13.1.1 融入知识创造过程的知识服务

e-Science 环境下的知识服务将是融入知识创造过程的服务，其目标是帮助科研人员更有效的完成知识发现的任务。知识发现的过程是知识的积累、转化和再生的过程，如图 13-1 所示。这是显性知识和隐性知识相互转化所构成的知识链的循环过程，图中的箭头表明了这种知识链的循环，因此知识服务是贯穿于知识链全过程的服务，其作用是加速知识链的循环速度，缩短循环时间，从而加速知识转化和再生的速度，以加速知识的创造。要实现这一目标，从根本上说要建立起基于知识链的服务链。

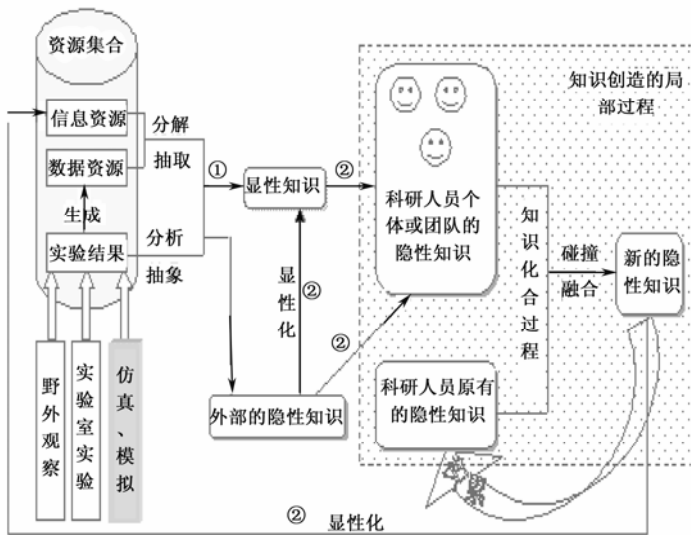


图 13-1 e-Science 环境下知识创造的过程

在 e-Science 环境下，知识创造的过程发生了两个方面的变化。①由于先进技术和理念的注入，知识创造的局部过程有了一些改变（如图中的灰色部分）；②各个环节的实现方法、承担者也可能发生改变，科研人员不必在每个环节上都投入大量的时间和精力，文献情报工作将在一些环节上发挥更大的作用。

具体分析 e-Science 环境下的知识创造过程，可以把这一过程划分为核心过程（图 13-1）中的虚线部分）和非核心过程，其中的核心过程是知识的质变过程，是科研人员创造性思维的过程，这个过程是别人无法替代的。而非核心过程则更多的是知识管理过程，是对知识的获取、组织、转化（显性知识与隐性知识之间）的过程，这一过程在一定程度上可以由文献情报工作者代替科研人员完成。

图 13-1 中①代表知识获取过程，这是一个从文献、数据等资源中抽取知识的过程。e-Science 环境下，资源的类型将会更加丰富，数量更加庞大，质量参差不齐，分布更加广泛，因此这一过程的工作量将会剧增。传统的由科研人员通过分析这些资源获得知识的做法将越来越不适应 e-Science 环境下科学研究活动的需求，需要寻找更便捷的知识获取途径，开发有效的知识获取工具，而这些活动都可以从科学研究的核心工作中剥离出来，由文献情报服务来实现。

图 13-1 中②代表知识转化过程。显性知识与隐性知识之间的四种类型的转化将交织着发生在四个环节上：

- （1）科研人员将由①获取的显性知识转化为个人或团队的隐性知识；
- （2）科研人员将外部环境中的隐性知识转化为个人或团队的隐性知识；
- （3）组织外部的隐性知识显性化形成外部的显性知识；
- （4）科研人员个体或科研团队的隐性知识的显性化，这种隐性知识是把吸收的知识与个体的隐性知识结合后，激发形成的新的隐性知识。

在 e-Science 环境下，一方面，从图 13-1 可以看到，知识转化所在的环节或者在核心过程之外，或者是核心过程与非核心过程的交互，这说明该部分活动是可以渗入的。另一方面，由于知识转化的速度将严重影响知识创造的效率，而 e-Science 环境下知识生产的数量和速度的急速递增又会制约知识转化的速度，科研人员将越来越难于兼顾这部分工作，因此这部分工作也需要文献情报人员的参与和帮助。

知识服务是以资源建设为基础的信息服务的高级阶段，e-Science 环境下的知识服务可以从以下几个方面展开。

（1）知识提供服务

知识提供服务是最直接的知识服务，是文献情报工作对知识获取过程的介入。

e-Science 环境下，提供知识的形式主要有以下两种：

- 直接知识提供。即按照科研人员的需求，通过对大量原始文献、数据等资源的汇总、分析、统计，挖掘、萃取知识，形成直接面向用户问题的答案，帮助科研人员解决科研过程中的问题，或者代替科研人员完成一部分知识准备和预研工作；
- 间接知识提供。即对原始文献、数据等资源进行分解、链接，形成知识元，向用户提供知识元及知识元“组装”工具，用户可以按照自己的需求，利用这些组装工具动态生成知识。由于不同用户的需求不同，利用这种工具可以由相同的知识元形成不同的满足用户个性化需求的知识体系。

简言之，知识提供服务是利用服务者所掌握和能够获取的各种资源为用户提供知识单元和知识集成服务。

（2）知识积累服务

这是对获取的或转化后的知识的记录和保存，承担起知识长期保存的功能。对知识的记忆和传承一直是文献情报工作者的责任，在 e-Science 环境下，这种责任的实现将从对知识载体的积累转向对知识的积累，通过对知识的提取、收集、组织和保存，使它在未来需要的时候能够被有效的利用。

知识积累包括两个方面：一方面是对资源中所蕴涵的知识的积累，即对显性知识的积累；另一方面是对科研人员头脑中所储存的知识的提取和积累，即对隐性知识的积累，即利用服务者所掌握的各种知识组织和管理技能，以用户在知识创造过程中生产的知识作为资源为用户服务。

（3）知识助理服务

知识助理服务是一种个性化的知识服务，是帮助科研用户实现知识管理的有效方式。知识管理是一种手段，一个过程，目的是通过对个体或一个动态团队所拥有的知识和外部知识的管理和利用，以提高个体或组织创造价值的 ability。e-Science 环境下，知识生产的速度和数量要求必须建立起高效的知识处理和管理机制，知识助理将帮助用户进行全方位的知识管理，提供知识管理的工具和方法，帮助用户建立起知识管理的环境并进行日常的知识管理工作，实现外部知识（系统）与用户的个体知识（系统）、团队知识（系统）的融合，以实现更高效的利用知识和创造知识的目的。

（4）促进知识价值的实现

科学研究所创造的知识的价值通常是通过科研成果的价值来体现的，而科研成果的价值则需要通过其应用来实现；同时，科研成果的应用会催生更多的科研需求的产生，从而使整个科学研究过程得以长期的持续。

e-Science 环境的建立，一方面使科研手段发生了革命性变革，从而引发科研成果数量的剧增；同时，由于对科研过程和中间产品的有效记录，使得各种形式的科研副产品不断增多，可以实际应用的不仅仅是科研的最终成果，还包括科研过程中产生的副产品，而这些副产品的应用市场和价值与最终的科研成果可能有着很大的不同。另一方面，科研人员将更加专注于科研的核心过程，而无暇于所创造的知识产品的应用；同时由于对市场的了解不够深入，有时也意识不到科研成果的市场前景和价值，市场的动态变化更使得新的商机随时涌现，这些可能在课题的可行性调研时还没有显现。因此，科研人员需要文献情报服务的帮助来促进知识价值的实现，搭建科研成果和市场应用之间的桥梁。文献情报服务可以通过建立虚拟的技术市场，提供新产品、新技术、科技政策、科技法规等信息，促进供需之间的信息交流与传递；通过对知识的捕获和传播扩大科研成果的知名度和影响面；通过对科研过程和成果的深入了解和市场的洞悉去挖掘新的应用市场和前景。

13.1.2 隐性知识的管理和利用

对隐性知识的管理是知识管理的一个重要方面。尽管知识管理还包括其他许多内容，但这里不准备对知识管理做全面的讨论，而是鉴于隐性知识对科学研究的重要性，对其进行重点的讨论。

按照波兰尼的知识分类理论，知识具有显性和隐性之分。显性知识是指由文字、图表、声音、图像等符号化形式表达和记录，外化到纸、磁介质、光介质等载体上，独立于人脑而传播的知识；相应的，隐性知识则是指难以用文字等形式记录，难以外化到载体上传播的知识，包括技术诀窍、技能技巧、经验、阅历、判断力、洞察力等。隐性知识通常以个人经验、印象、感悟、团队默契、技术窍门等形式存在于人际沟通、群体交往及社会文化的延续过程中，其数量和重要性要远远大于显性知识。一般认为，在组织的知识结构中，80%以上的知识是隐性知识。如果把知识体系比喻为一座冰山的话，那么显性知识只是冰山露出水面的一小部分，而隐性知识则是冰山隐藏在水下的巨大部分，它们不仅在数量上占绝对优势，而且在重要性方面也很突出，就像冰山下部所起的支撑作用一样。

科学研究的过程实质上是知识创造的过程，而知识创造的过程实质上是隐性知识和显性知识二者之间互相作用、互相转化的过程。按照野中郁次郎的知识转化模式，这两种知识之间存在着四种转化过程，四种过程构成了一个有机的知识转化整体，以往文献情报工作介入较多的是汇总组合过程，即显性知识→显性知识的转化，对这一部分也较易于把握，而对其他三种与隐性知识相关的知识转化过程则关注和介入较少。

分析与隐性知识相关的三种转化过程，隐性知识→隐性知识的转化一般通过授体与受体之间直接的交流实现，因此需要有通畅的交流渠道和良好的交流秩序作支撑；显性知识→隐性知识的转化是吸收者的智力因素决定的，是受中介因素干扰较小的转化过程；而隐性知识的显性化则是科研成果的一种体现，同时也是实现隐性知识在更大范围内共享的唯一途径。

1. e-Science环境下隐性知识的特点

(1) 隐性知识在科学研究中的重要性

首先，隐性知识是科研成功的决定性因素。松下电器公司发明面包机的过程生动地体现了这一点，他们通过观察、总结和试验成功地模拟了大阪国际饭店首席面包师的拉面技术，通过对这一隐性知识的成功再现取得了面包机的成功。面包师的拉面技术这项隐性知识成为面包机研制成功的决定性因素。

朱克曼的研究更是有力的证明了这一点。她在 20 世纪 70 年代研究了杰出科学家之间的关系，结果表明：截止到 1972 年，92 名美国诺贝尔奖获得者中有 48 名曾经作为老诺贝尔奖获得者的学生、博士后或年轻的同事。通过对许多杰出科学家系

列的研究表明了诺贝尔奖获得者通过师徒关系在不同代际之间的延续,这些未来的诺贝尔奖获得者在作为“徒弟”的时候,从“师傅”那里主要学到的不是显性知识,而是诸如工作标准和思维模式等不能编纂整理的思维和工作方法等隐性知识。隐性知识在科学研究中的重要性由此可见一斑。

其次,隐性知识的显性化程度决定了科研成果的价值大小和隐性知识价值实现的大小。隐性知识显性化的结果是科研过程中所创造的知识的具体体现,在一定程度上代表了科学研究的成果,因此转化的程度决定了科学研究所创造的知识的价值。另一方面,转化的程度还决定了隐性知识能够被共享和利用的程度,因为直接交流所传播的隐性知识的范围是有限的,隐性知识只有在显性化之后,才能够在更大的范围被更多的人享用,因此显性化程度决定了隐性知识的再生价值。

(2) e-Science 环境加剧了隐性知识难以管理的特点

隐性知识本身具有难以管理的特性,这主要体现在它难以显性化方面,原因在于:① 隐性知识是一种感悟、一种直觉,而这种感悟和直觉本身具有难以表达、难以传递的特点;② 由于其拥有者个人的知识水平和表达能力有限,限制了对隐性知识的总结和表述;③ 缺乏有效的工具和环境促进隐性知识的传播、交流和显性化;④ 缺乏有效的体制和机制保障隐性知识拥有者的权利,缺乏有效的激励机制激发拥有者贡献、表达隐性知识的热情。

e-Science 环境加剧了隐性知识难以管理的特点。在 e-Science 环境下,科研活动将按照面向问题和任务形成跨地区、跨行业、跨机构的虚拟组织,这种目标团队的广泛性、虚拟性、动态性和临时性将使其产生的知识和诀窍等隐性知识比以往更难以保存、传递和再利用。

(3) e-Science 环境下文献情报工作的转移

传统的文献情报服务关注的焦点是显性知识,确切地说,是显性知识的载体——印刷型和电子版的各种资源,围绕着这些资源的收集、组织、整序、检索和利用展开工作。而当对显性知识的管理趋于明朗化、规范化、常规化之后,能否有效地挖掘和管理隐性知识将成为竞争的焦点和培育核心竞争力的关键。

e-Science 环境下,知识服务将成为文献情报工作的核心竞争力,而要做好知识服务,除了关注显性知识之外,对在科学研究中起关键作用的隐性知识的管理将是文献情报工作融入 e-Science 环境的一个突破点。近年来,已经有深入到资源中提取知识、提供知识服务的趋势,但知识服务所关注的焦点仍旧在显性知识的挖掘方面,即如何从大量资源中获取知识的问题,而隐性知识在科学研究中的重要性,文献情报工作对知识创造过程和科学研究过程的渗入都要求文献情报工作更多的关注对隐性知识的管理。

2. e-Science环境下管理隐性知识的策略

(1) 为隐性知识创造“说话”的方式

e-Science 环境下文献情报工作对隐性知识的管理应首先集中在隐性知识显性化方面。文献情报工作者通过融入知识创造的过程,将可以承担起帮助科研人员表述其隐性知识的作用,弥补科研人员不善于或无暇总结、表达隐性知识的缺欠,推动隐性知识的显性化。

e-Science 为隐性知识的显性化提供了更强大、更方便的工具,模拟和仿真技术能够更真实、更准确、更直观地重现科研人员头脑中的景象,在一定程度上解决了隐性知识难以用语言描述的问题。文献情报工作可以利用 e-Science 所提供的技术和工具,根据自身对文献、信息、乃至显性知识管理的理论和经验,研究促进隐性知识显性化的工具,促进隐性知识的显性化。例如,利用多媒体、超媒体等多种形式的载体,可以更形象的表达隐性知识。

(2) 为隐性知识的转化创造交流的环境

知识交流是实现隐性知识与隐性知识之间转化以及隐性知识显性化的有效方式之一。e-Science 环境下,知识交流将日趋复杂化,它将与数据交流、信息交流交织在一起,与科研过程交融在一起,知识交流的平台将发生转化,数字化科研环境的建立也同时实现了知识交流平台的数字化、网络化。

文献情报服务通过对构建新型知识交流环境的参与,对知识交流的运转体制和机制的约束、控制、协调和管理,对知识交流的过程和成果的规范和管理,推动开放存取(open access)、OAI、e-print 等新的学术交流手段不断发展,从而推动隐性知识的直接传递和显性化。

13.1.3 分布式资源体系的建设和管理

1. 资源建设的目标

e-Science 环境下,为了实现对资源的共享,首先必须有效地组织和管理资源。为此,必须建立一个能够充分揭示资源的,实体分布、虚拟集成的资源体系,通过资源的耦合与共享形成一个良好的资源空间。在 e-Science 环境下,资源是分布的,而且是有控制的分布的,为此,需要建立一种在分布的资源体系下协调使用的保障机制,以确保资源是分布而不是分割的,资源在地理上是分布的,可以分布在不同区域、不同机构、甚至就在科研人员那里,但在逻辑上是一个集成的整体,确保科研人员在需要时可以从任意资源节点获取资源。

在此基础上,要建立有效的获取途径。揭示仅仅是一个手段,是获取过程的一个环节,而获取则是更高层面上的问题,是资源建设的最终目标。有效的获取途径就是要降低用户获取资源的成本,提高用户的资源满足率。

2. 以灰色资源的管理和利用为重点的信息资源建设

资源建设是一个涉及面很广的概念,不同类型的资源有不同的特点和建设方式,这里仅对信息资源中灰色资源的建设详加阐述。

(1) e-Science 环境下对灰色资源的界定

在传统的文献情报服务中,人们习惯于按照出版发行方式将文献划分为白色文献、灰色文献、黑色文献,套用这种对文献的划分方法,也可将信息资源划分为白色资源、灰色资源、黑色资源。

- 白色资源:指通过公开出版发行渠道出版发行的信息产品或资源。
- 灰色资源:指未进入公开出版发行渠道的信息产品或资源,包括传统的科技报告、会议资源、学位论文、标准、专利,以及网络化环境下的聊天日志、视频会议记录、科研过程记录、开放源码的软件产品等。简言之,在 e-Science 环境下,灰色资源的含义将超越对文献的界定,包括整个科研过程本身及科研过程中产生的一切副产品中不能或没有正式出版的资源。
- 黑色资源:指由于保密而无法公开出版发行的信息产品或资源。严格地说,它也是灰色资源的一种,但由于密级限制,使它受到法律的保护,在合法的手段下不可能通过改善检索途径等方法获得。

(2) e-Science 环境下强调建设灰色资源的原因

e-Science 环境下,资源建设的重点将转向灰色资源的建设,这是因为:

- 灰色资源在科学研究中的重要性与日俱增:据美国科学基金会及日本国家统计局调查,世界上 90% 以上的新技术通过灰色文献中的专利文献公诸于世,95% 的产品技术情报来源于灰色文献中的产品资料。在 e-Science 环境下,由于非正式学术交流的频繁,灰色资源的数量将会迅速增多,它将成为 e-Science 环境下资源构成的重要组成部分,而且由于它所传递的信息新颖、及时,它将成为科研人员获取信息的主要渠道。
- 灰色资源难以获取和管理的特性日益加剧:较之于白色资源,灰色资源的获取更加困难,究其原因,主要有以下三个方面:①由于没有进入公开的发行渠道和正规的流通渠道,获取不易;②对它的揭示不够,检索不易;③囿于版权或保密等原因,获取原文不易。e-Science 环境下交流范围的广泛性、交流手段的多样性将使得这一状况加剧。
- 白色资源的建设已趋成熟:从目前的现状分析,白色资源的建设正在不断完善和日趋成熟。各种文献情报机构通过协调采购、联合采购、集团采购等多种形式建设资源,通过跨库检索、学科门户等多种方式整合资源,使这部分资源的覆盖率和满足率不断提高。相比之下,对灰色资源的建设还比较薄弱,其满足率尚不能令人满意。

鉴于灰色资源的重要性和它的难以获取性以及目前资源建设的状况,在

e-Science 环境下, 资源建设应该更多地倾向于灰色资源的建设。分析灰色资源获取困难的三个原因可以发现, 亟须建立有效的资源揭示机制和通畅的获取途径, 保障对灰色资源的全面揭示和有效获取, 提高灰色资源的满足率。

(3) e-Science 环境下灰色资源建设的主要内容

- 完整的科研过程记录: 完整的科研过程记录包括对整个科研过程(从科研思想的萌发到科研成果的应用)的系统、详尽的记录, 不仅有成功过程的记载, 更有对失败过程的记录。事实上, 每一次成功的背后可能会有无数次失败的尝试, 详细地记录这些失败的实验过程, 如使用的材料、方法、工具、条件以及失败的原因分析、所采取的补救或修正措施等, 可以使后来者避免同样的错误, 或许后来者还可以通过对使用过程的分析, 对材料、方法、工具、条件的分析, 结合自己的隐性知识, 另辟蹊径, 发现新的组合, 从而导致新物质的产生, 新方法和工具的创建, 实现创新。传统的科研活动中, 科研记录制度不规范、手工记录的方式单一、烦琐、落后, 使得对科研过程的详细记录难以持续; 而在 e-Science 环境下, 可以利用各种先进的技术和工具, 有目的、有选择的自动记录科研过程。
- 灰色资源的系统化、集成化、共享化: 灰色资源通常零散的分布在科学研究的过程之中, 处于一种零乱、无序的状态, 需要按照不同的分类对其进行组织和揭示, 利用不同的元数据体系标识这些资源, 使之可以通过不同的入口获取和不同的渠道传播, 同时需要对不同的组织体系进行有效的集成, 使之可以通过统一的接口呈现给用户。
- 对灰色资源的全面揭示: 从揭示的层面来说, 不仅要求对资源的内容特征进行揭示, 还要求对资源的形式特征进行深入揭示, 如作者、出版者、保存者等相关资料的详细记录和描述。从揭示的方式来说, 对资源的揭示应该是动态的, 尤其是对形式特征的揭示, 要能够使资源的组织系统与其他系统(如出版系统、人力资源系统)横向结合, 动态地跟踪、搜索形式特征的变化, 并及时的修改本系统中的相应指标。

13.1.4 科学数据的管理

2002 年, 科技部部长徐冠华院士在主题为“中国科学数据共享”的香山会议上指出: 当代科学技术的发展明显呈现出大科学、定量化、重过程研究等特点, 越来越依赖于系统地、高可信度的基本科学数据及其衍生的数据产品。

1. 科学数据的管理——e-Science带来的新问题

e-Science 衍生了数据洪流。e-Science 环境中, 将会有大量高性能的仪器设备投入使用, 这些新型仪器设备以及现有设备的协同工作将使科学数据的生产速度和效

率发生质的变化，数据量级将由 GB 到 TB、PB 级。例如，将于 2007 年投入使用的 LHC（大型强子对撞机），每年将产生 1~100TB 的数据，这相当于二千万张 CD 的数据量。在粒子物理学和天文学领域，由于新型实验设备的联机使用，预计在今后 10 年里每年将产生 100TB~10PB 的数据，生物信息学、医学等其他领域也将出现类似的情形。

科学数据在科学研究中的重要性不断提高。用不同的工具或方法、在不同的历史阶段对同一组科学数据进行分析可以产生不同的结果；越来越多的科学研究不再从头开始，而是建立在对已有数据的重新认识、组织、解析、分析和利用的基础上；科学数据越来越成为证实或驳斥一种观点或想法的最有力、最直接、最客观的证据。种种迹象表明：科学数据在科学研究中的重要性正日益提高。

鉴于此，需要收集、整理、存储、组织、传递、保存这些科学数据，以便它们在适当的时候可以被适当的人重新利用。要实现对数据资源的有效管理，除了依靠先进的技术设备以外，还需要对这些科学数据进行有效的存储、组织、检索，因此，需要有统一的、有利于互操作的元数据标准规范来描述科学数据，合理的知识组织体系、灵活方便的数据存储格式来存储、组织数据，可靠的通信机制来传递数据，适当的共享机制来共享数据，有效的长期保存机制来保存数据。而不同类型的科学数据具有不同的属性，其管理的方式和工具也会不尽相同。

2. e-Science环境下对科学数据的管理

e-Science 环境下对科学数据的管理包括如下几个方面。

- 收集科学数据。由各种仪器设备产生的科学数据是分布在全球各地、各个机构中的，对这些数据需要进行有效的收集。可以是物理上的集中，建立数据存储和处理中心，也可以是逻辑上的集成，建立一种链接机制，无论何种方式，收集的最终目的是提供充分的揭示和获取机制，实现数据的共享和利用。
- 保存与共享科学数据。为了实现数据的长期可用，必须对其进行永久的保存。e-Science 环境下，数字化的数据生产和传递将有利于它的存储和转换，也使得对它的保存和重用成为可能。例如，在动物和微生物基因组研究方面，美国科学家首次绘制出了黑猩猩的基因组序列草图，所获得的全部数据存入公共数据库，免费供各国科学家调用。科学家们通过比较分析发现，黑猩猩与人类的基因组之间存在着关键性差异，说明二者的共同祖先在五百万年前分开之后，自然选择压力的差异导致了两种生物进化过程的不同。正是这种对数据的重用推进了科学研究的进程。在数据的保存与共享中，需要考虑数据的组织、长期保存的策略、数据的迁移、检索等方面的问题。
- 分析与处理科学数据。科学数据中蕴涵着大量的事实和规律，而这种事实和规律的揭示是通过对大量数据的分析和处理实现的。因此，需要建立科学数据的分析和处理工具，以便能够从不同的需要出发对数据集进行各种角度的

统计、计算、分析和处理，帮助科研人员进行数据的预处理和后处理工作。

13.1.5 构建开放的数字化网络化学术交流体系

1. 传统的学术交流体系及其在科学研究与知识创造中的不足

按照门泽尔（Menzel）和米哈依诺夫（A.H.Mikhailov）对“正式交流过程”和“非正式交流过程”的划分，传统的学术交流体系是以正式交流过程为主导的，它以文献等资源的出版发行为标志，这种交流体系在科学研究的特定历史时期曾发挥了巨大作用。但是，随着数字科研时代的到来，这种交流体系的局限性也逐渐突显出来，体现在：

（1）仅仅记录了科学研究的结果，但无法记录科学研究的过程。

（2）价格壁垒：出版商的垄断经营，利益驱动使得出版物价格飞速上涨，限制了对其的访问和使用。

（3）出版周期长，使得科研成果的发表具有严重的滞后性，有限的版面无法容纳越来越多的科研成果，影响了科研成果的及时传播。

（4）同行评议的局限：由于带有一定的主观性，有时会限制新思想的传播，而且还会有偏见存在。

（5）交流过程垂直、僵化：学术交流单元（如一篇文章）在学术交流体系中的流动是单一、线性的。文章在投稿之后，经过同行评议、修改，到最终的出版，每个环节都只能由一个学术交流中介完成，无法由多个中介同时进行。

（6）版权障碍：在正式出版体系中，出版社在出版作品时一般要求作者放弃著作权。例如，美国化学学会声明：“作者可以向不多于 50 个人发布或传递他们的论文，也可以在他们的个人主页上发布论文的题目、摘要和图表。”版权的转移使作者丧失了对自己作品的处置权，按照上述声明，如果作者向超过 50 个人提供了他的作品，就构成了侵权，这种版权保护限制了作者对作品的灵活处理，限制了对作品的使用，甚至是合理的使用。

2. 文献情报工作在传统的学术交流体系中的角色和作用

文献情报工作在传统学术交流体系中的作用是优化输出，如图 13-2 所示，即对大量无序的信息流进行收集、整序，以有序的、方便读者利用的方式提供给读者。具体说来，其功能体现在如下两个方面：

（1）保存功能：对各种类型学术资源的收集和保存，以确保在未来某个时刻对学术资源的存取和利用；

（2）传播中介：作为作者与读者之间的中介之一（还有出版等其他中介），承担着对交流单元的采集、整理（分类、描述、编制检索工具等）、提供服务的功能。

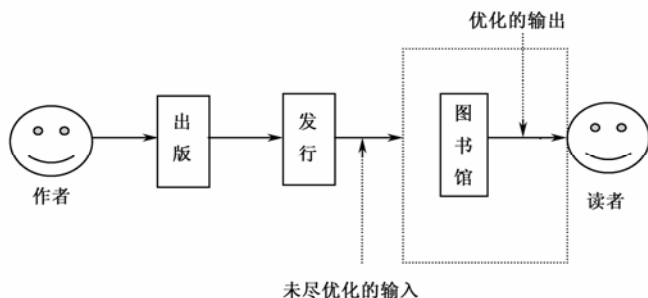


图 13-2 文献情报工作在传统的学术交流中的作用

3. e-Science环境下学术交流体系的特点

e-Science 环境下的学术交流体系将是一种开放的数字化网络化学术交流体系，呈现出如下新的特征。

（1）交流的非中介性

e-Science 环境下，科研活动是基于网格实现的，技术的进步使得交流的时空障碍逐渐消失，科研人员将有更多的条件绕过出版发行商等中介机构，交流往往在交流主体之间直接发生，不需要中介的参与。这种非中介性使得交流的发生更加灵活，受到的限制更少。

（2）交流单元构成的复合性和细化

一方面，正式出版物将不再是构成交流单元的主要内容，数据集、软件、动态知识描述、模拟结果等都可以作为交流单元参与交流。另一方面，在传统的学术交流体系中，交流单元以出版物形式体现，用户只能购买一本书或一本刊，而不能购买书中的一个片段（如某一章、一节、一个要点或一句话），而在新的交流体系中，这些片段将作为独立的交流单元参与交流和传递。

（3）交流的多向性

在传统的学术交流体系中，交流是单向进行的。例如，对于某一篇文章，读者对文章的观点、评论要想进入交流系统，必须经过“投稿→同行评议→出版”这样一个交流链的循环，实际上是启动了一条新的交流链，这使得交流难以实现互动。而在 e-Science 环境下，交流将更多地体现互动性，交流双方可以对交流单元进行讨论、研究，这个过程可以互动地实现并自动记录。

（4）原生数字化资源增多

由于交流过程的数字化，伴随着交流过程产生的副产品也更多的以数字化的形式存在，缺少对应的印本资源。这使得对资源的长期保存问题更为重要。

（5）具有一定的局限性

表现在：①非正式交流的随意性强，用户可以自由的发布信息，缺少严格的审

核机制,缺乏有效的控制机制,其质量、时效、稳定性难以得到有效的保证;②非正式交流的方式多样,电子邮件、BBS、讨论组、网络会议、信息共享室、个人主页……这就产生了大量形态各异、质量参差不齐的资源,良莠混杂使得对这些资源的筛选和甄别变得异常困难,如何有效存储和利用这些资源的问题将日趋尖锐;③非正式交流的组织性差,许多非正式交流是科研人员自发形成的,这使得交流的过程和结果极不规范,不易记录和保存。

4. 文献情报服务在e-Science环境下学术交流过程中的作用

如前所述,e-Science 环境下科学交流的模式将发生嬗变,开放的数字化网络化学术交流体系将逐渐形成。这种学术交流体系的重构,不仅需要硬件和技术的支持,还需要解决一系列的相关问题,如学术交流系统的互操作问题、工作流问题、服务共享问题、信息建模问题等。作为文献情报工作者,尽管不需要从技术上实现这种学术交流平台,但在技术平台构建起来之后,如何使科学交流的过程能够顺畅的在这个平台上实现则是文献情报服务机构的责任。因此,需要对交流的主体、交流的行为、交流的过程、交流的结果进行必要的引导和规范,以保证交流过程中闪现的信息、知识、思想火花能够被有效地记录和重用。文献情报服务在 e-Science 环境下学术交流过程中的作用体现在以下两个方面。

(1) 交流过程的记录和共享

交流过程是一个动态的、交互的过程,具有双向性,它实质上是交流主体之间就某一问题进行深入讨论的过程,属于科学研究的一个阶段。在传统的学术交流系统中,这一过程是难以记录和再现的。例如,在期刊论文的发表过程中,读者看到的只是最终发表的文章,文章的若干修改版对读者来说是不可见的,同行评议结果对读者来说是不可见的,其他读者对文章的评注也难以进入交流领域,这使得读者不能跟踪作者思想的发展和变化,无法全面了解他人的观点和思想。e-Science 环境下,科研人员之间的交流将更多的通过网络完成,借助于计算机强大的记忆和存储功能、后台的自动记录操作,有可能完整地记录和重现这种交流过程,而这种交流过程中所隐含的信息、知识则有可能被抽取、组织和共享。

(2) 交流过程中产生的各种类型资源的记录、存储、共享和重用

e-Science 环境下,一方面学术交流各个环节的结果将呈现类型多样化、数量巨型化、载体数字化、生成随意化的趋势,使得交流单元的组成多样化;另一方面交流单元进入交流系统的时间和位置节点呈现多样化的趋势,这使得传统的资源组织方法不再适用,需要新的工具和方法来对这些交流单元进行发现、组织、质量控制、检索,以实现它们的共享和利用。

13.1.6 以网络信息意识的强化为重点的用户信息素养的培育

用户的信息素养包括三个方面：信息意识、信息能力和信息道德。信息意识是一种思维力，是用户积极思考、敏锐洞察、深入挖掘、准确表现个体信息需求的能力；而信息能力则是一种行动力，是用户技巧地利用各种检索工具有效地满足个体信息需求的能力。具体说来，信息意识包括两个方面，一方面是科研人员对外界信息的意识能力，他应能广泛地接受各种信息，敏锐地意识并提取出有价值的信息以应用于科研实践；另一方面是对自身信息需求的意识并准确表达这种需求的能力。信息能力的组成则是多方面的，它包括：①熟练使用各种检索工具的技能；②对检索结果的识别、整序、组织、过滤、析取、重用的能力；③根据个体的需求建立个性化知识体系的能力，即知识组织与管理的能力。信息道德是指在整个信息活动中，调节信息创造者、信息服务者、信息使用者之间相互关系的行为规范的总和，它是约束信息生产、传递、消费过程的一种道德规范，是个人的道德品质在科学研究过程中的体现。

1. e-Science环境下的信息素养分析

在 e-Science 环境下，“科学研究的有效性极大地依赖人们的信息生存能力……知识的发现、利用和创造不再仅仅依赖‘好的大脑’和‘坚韧不拔的精神’，而是依赖于人们能否有效地利用相关的信息技术、工具和系统。”，用户的信息素养将成为成功的知识创造的主导因素之一。

信息素养与信息需求密切相关，按照美国学者 Nancy Lemon 的观点，用户的信息需求可以归纳为以下两点：①用户知道信息的存在，或者不知道信息的存在；②用户知道所需信息具体在哪里，或者不知道信息在哪里。

从认知学角度分析，用户需求有两种存在状态：现实的与潜在的（显性的与隐性的）。前者指用户已经表达出来的或已经意识到但尚未表达出来的需求，而后者则是指潜在的、尚未被用户所意识到、所认知的需求。从认知学的角度对 Nancy Lemon 的用户需求框架进行修正，可以得到 e-Science 环境下新的用户需求框架模型。这个模型也可以归纳为两点：①用户知道需求的存在，或者不知道需求的存在；②用户知道如何满足个体的需求，或者不知道如何满足个体的需求。

审视 e-Science 环境下的用户需求框架模型可以发现，用户的信息素养在这个模型中得到充分体现。按照所归纳的两点，其中“用户知道需求的存在，或者不知道需求的存在”是用户的信息意识问题；而“用户知道如何满足个体的需求，或者不知道如何满足个体的需求”则是信息能力的问题。

仔细分析用户信息素养与信息需求的关系可以发现，如果用户停留在区域①中，说明他们既知道信息需求的存在，也知道如何满足这种需求，他们的信息素养是可

以满足现实的信息需求的；如果用户处在区域②中，则表明他们虽然意识到信息需求的存在，但还不知道如何满足这种需求，造成这种结果的原因有用户主观方面的原因和外界环境客观的原因，用户的主观原因是用户的信息能力不足造成的，而客观原因主要是缺乏通畅的信息传递渠道所致；如果用户处在区域③中，则表明他们还没有意识到个体的信息需求，这是用户的信息意识不足造成的，而信息意识和观念是主观的、他人无法替代的，无论图书馆的服务多么完善，它不能代替科研人员从事科研创造，因而它不能代替科研人员的主观意识。

2. e-Science环境下信息素养的培育

e-Science 环境下，科研用户信息素养的培育与普通用户相比有许多特殊的方面。一方面，现代信息技术的应用和信息服务的不断深入使得信息获取的壁垒不断降低；另一方面，多年来普遍受到重视的用户信息素养的培育主要集中在用户网络检索能力的培训方面，这使得用户的信息检索能力不断提高；最后，分工的细化使得信息获取这种非核心业务逐渐从科研用户的核心工作中剥离出来，不再是科研用户关注的焦点。因此，e-Science 环境下科研用户信息素养的培育应该从以下三个方面来考虑。

① 信息素养培育的重点和焦点转向网络信息意识的培养和强化

信息意识的培养应着眼于帮助用户提高他们的信息敏感度。文献情报工作可以代替科研人员完成信息的检索、整序、过滤、组织、以至知识的析取，但不能代替科研人员进行思维，而只能作为外因刺激不断激发用户的信息意识，促进潜在需求向现实需求的转化，而要实现真正的转化，还需要科研用户对需求的最终认识和认可。

② 信息能力的培养应集中在知识管理能力和知识转化能力

这包括诸如有效地组织知识，熟练地使用知识挖掘工具，将个体生产的知识与外部知识的有机融合等方面。面对相同的资源，不同的人会有不同的理解和收获，图书馆不能期望对用户的每一个问题都提供知识型的回答，用户的需求也并非如此。因此，必要的信息能力仍将是进行科研活动的必备条件之一。而信息能力的培养向知识管理能力方向的转移源自于 e-Science 环境下信息获取的便易性，丰富的数据、信息与贫乏的知识之间的矛盾。用户信息获取的主要障碍已经不是距离而是如何选择信息、如何使用各种工具从信息中抽取对解决问题有用的知识、如何管理内外部知识的问题。当知识提供给用户时，当知识挖掘、知识转化工具提供给用户时，能否有效的利用这些知识和工具将成为科研成功的关键。

③ 建立一个诚信社会

科研道德建设是一个日渐沉重的话题。之所以这样说，是因为目前学术界的诚信危机。从图书馆书籍报刊中的一幅幅天窗，到一再曝光的论文抄袭、剽窃事件，无不反映了正在不断下降的信息道德素质。但是，科研活动是公平、客观的，不应有一丝虚假，信息道德左右了一个人提供信息、使用信息的态度，也间接影响着他的科研态度。在 e-Science 环境下，开放的科学交流体系的建立除了要靠法律手段、

技术手段的规范外，还需要科研人员的自我约束，只有在一个诚信的科研环境中，科研人员才会有热情、有勇气、愿意贡献自己的知识，开放存取的实现才能够畅通无阻。

总的说来，在 e-Science 环境下，用户信息素养的培育重点将转向强化信息意识的培养，以消除信息盲区，同时提高用户的知识管理能力和知识转化能力，促进知识共享，减少知识流失。

13.2 e-Science环境下文献情报机构服务的实现模式

13.2.1 服务理念的提升——从“以需求拉动服务”到“以服务激发需求”

20 世纪后半叶以来，图书馆的服务重心开始从“书本位”向“人本位”转变，文献情报机构的服务理念也随之向“以用户为中心”转移，“满足用户需求”逐渐成为文献情报工作的服务宗旨和目标。在这一目标的拉动下，文献情报工作在资源建设、服务提供、系统发展、平台建设等诸多方面始终为如何最大化的满足用户需求而努力探索着。

传统的文献情报服务更多的是当用户已知需求但不知道如何满足这个需求的时候，文献情报服务通过有效的获取途径和工具帮助用户满足这个需求。

但是，用户的信息需求不仅仅有现实的，还有很大一部分是潜在的。潜在需求向现实需求的转化有两种动因：一种是用户的内因作用，用户在求解问题的过程中可能会产生新的问题，在学习过程中自然积累的知识达到一定程度，与用户原有的知识化合而发生质变时，也可能会引发新问题的产生，这些都会导致潜在需求的显性化；另一种是外因的作用，如一些突发事件的刺激、环境变化的诱导，这些外因会加速潜在需求向现实需求的转化，甚至能够创造用户需求。

e-Science 环境的形成给科研活动带来了巨大的变革，科学研究的重构，学科的分化、融合、衍生加剧，新问题大量涌现，科学研究进入一个新的不稳定期，这些将会引发更多的潜在需求的产生。因此，在 e-Science 环境下，文献情报服务不应该再停留在满足用户需求的阶段，而是需要进一步提升服务理念，从满足用户需求向激发用户需求发展，变“以需求拉动服务”为“以服务激发需求”。也就是说，文献情报机构的服务理念将在区域③中得到充分体现，不仅要满足用户现实的需要，更要进一步激发用户潜在的信息需求。文献情报服务需要扮演外因的角色，利用日志分析、跟踪用户创造知识的行为，掌握用户行为的变化，利用文献计量学方法、引文分析、内容分析、链接分析等方法掌握科学发展的趋向和变化，由此预测用户的需求，以此为依据提供服务，激发用户的潜在需求，进而满足这一需求。

文献情报服务不仅仅是激发潜在需求的外因，同时还是内因的刺激因素。通过不断强化用户的信息意识可以使用户从主观上不断挖掘个体的需求，实现潜在需求

向现实需求的转化。

13.2.2 组织模式的分化——嵌入式和支持中心的双层模式

为科学研究服务的专门图书馆和科技文献情报机构可以进一步细分为如下两类：一类是隶属于科研机构等单位的文献情报部门：这类文献情报部门有一个带有“血缘”关系的母体，这种关系决定了其生存的意义和目标便是为母体机构的科学研究和技术开发服务。它们在经费供应、人员、编制等方面受制于母体机构，它们具有明确、专一的用户群，其需求面窄、度深、相对集中、学科属性相对单一。另一类是综合性专业图书馆、科技文献情报机构：这类文献情报机构一般作为一个独立的法人实体而存在。它们一般不是针对某个机构的用户服务的，而是面向某个区域、某个系统的科研用户提供服务的，而这些用户可能从事不同学科领域、专业方向的研究工作。

针对知识发现的过程，文献情报工作应该建立基于知识链的服务链来满足用户需求，实现自身价值。上述两种不同类型专业文献情报机构的特点和分工的不同使得这种服务链在 e-Science 环境下的实现具有不同的组织模式。概括起来，这应该是一种双层模式，是一种双向拓展、纵向贯通的模式，内层是支撑层，由综合性专业文献情报机构承担，而外层是实践层，由附属的专业文献情报机构承担，这个两层的组织和服务模式如图 13-3 所示。

1. 以嵌入式为核心的组织模式

文献情报工作嵌入到科研工作中，文献情报工作者嵌入到科研团队中，成为其组成部分，这是基层文献情报机构融入 e-Science 环境的主要方式。从项目的调研、论证、立项，到开展研究，到最终的结题验收、推广应用，提供全程的、持续的、深入的服务是未来科研活动对文献情报服务的基本要求。

文献情报服务嵌入到科研活动之中可以有多种形式，最常见的有如下三种。

(1) 过程嵌入

文献情报工作嵌入科研工作与科研过程之中包括两个方面的内容：① 对科研过程中生成的各种信息资源的收集与管理。以往这个过程是与科研过程、与资源的生产过程相分离的。最典型的要属学术论文的管理，许多科研机构是在年底统计时才比较全面地收集一年中本机构的科研人员发表的论文，而收集的结果也并不总是令人满意，对于那些非正式的资源就更加难以收集。数字化科研环境在对这些资源的共享和利用提出更强烈的要求的同时，也为有效的收集这些资源带来了便利条件，由于资源的生成是在数字化的科研环境中产生的，通过建立相应的规范、工具、方法，一方面可以通过一些程序在资源生成过程中在后台自动收集这些资源，无需人工干预，同时也可以使科研人员在科研过程中对生产的资源自觉的进行规范化的提交和存储，例如通过程序提醒科研人员，或者在没有完成提交时无法进行后续的工

作。② 服务过程与科研过程的融合。以上海有机化学所为例，它的信息中心包括一台服务器，负责从核磁共振实验室的核磁共振仪器中提取数据，对之进行计算并生成谱图，然后将这些谱图分发给科研人员，科研人员可以根据自身的需求选取、使用并对之进行处理。以往谱图的生成是由科研人员完成的，是科研活动中的一个环节，而在 e-Science 环境下，可以将这一环节从科研过程中剥离出来，借助于仪器设备的帮助，由文献情报工作者代替科研人员完成它，成功地实现了文献情报工作与科研工作的结合。

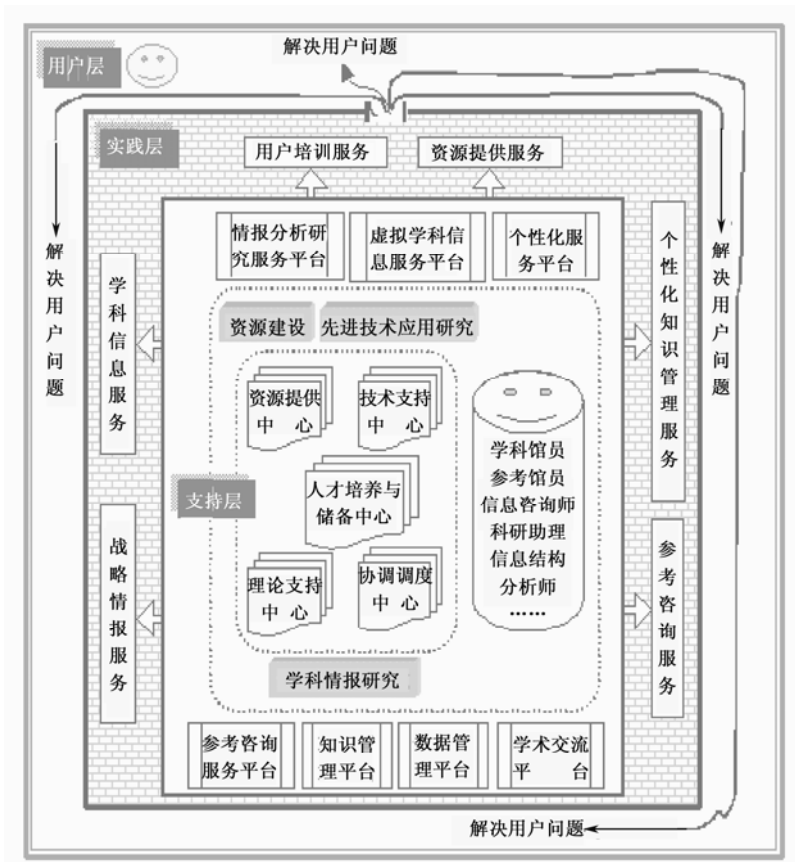


图 13-3 e-Science 环境下文献情报服务

(2) 人员嵌入

使文献情报工作者嵌入科研团队，成为科研人员的一分子。文献情报工作者嵌入科研过程之中，可以全面地了解科研的进程，动态跟踪科研的发展，可以有效地发挥两个作用：一是帮助科研人员及时解决科研过程中遇到的问题；二是通过直接参与科研过程，通过与科研人员的交流和观察帮助科研人员实现隐性知识的显性化。

(3) 平台嵌入

无论是过程的嵌入,还是人员的嵌入,在 e-Science 环境下主要是通过平台的嵌入来实现的。因为,在 e-Science 环境下,每一项服务的实施都是依托于一定的平台进行的,如资源的自动采集需要有相应的采集平台和软件的支持,只有这些平台真正与科研平台结合在一起,成为科研平台的一个组成部分,才能够为这些平台上的服务创造与科研活动结合的机会。因此,在开发和研制服务平台的时候,要充分考虑它与科研平台的衔接与融合。

2. 知识加工与服务支持中心模式

(1) 支持中心的两层含义

① 对实践层的支持。实践层要完成为科研服务的任务,仅仅依靠自身的力量是不够的,它还需要从技术上获得支持,资源上获得满足,人力上获得补充。而这些支持、满足和补充将是支持中心的责任。强调支持中心,就是要使这类文献情报机构清醒地认识到,它所服务的用户不仅仅包括科研人员,还包括基层文献情报工作者,很多时候,为科研用户服务的理念和行动需要通过基层文献情报工作者的实践和参与来实现。

② 对科研活动的支持。它包括两个方面:一是对所有的科研用户和科研活动提供间接支持;二是对重点用户提供直接支持。通过这种支持,一方面可以更好地满足科研用户的需求,另一方面也可以积累可供推广的经验。该支持中心的模式,并不是要抛弃嵌入式服务,当这类机构服务于科研用户的时候,嵌入式服务仍将是主要的服务方式。这类机构对科研活动的支持不能面面俱到,只能是在某些服务点上、对某类用户来实现。

(2) 支持中心的任务

资源提供中心:支持中心应该承担起资源建设的责任,通过建立广泛的资源获取渠道,不仅保障对白色资源的获取,还要保障对灰色资源的获取,不仅保障对文献型资源的获取,还要保障对事实型、数值型资源、多媒体资源、甚至仪器设备、计算资源、人力资源的获取和提供。

技术支持中心:e-Science 环境下的文献情报服务将是技术依赖型的,各种服务的提供都离不开技术的支持,基层文献情报部门往往囿于人力等原因而无法很好的开发服务所需的各种技术和平台,尤其是一些通用的平台。支持中心可以根据文献情报服务的需要开发各种可定制的通用或专用工具或平台,如数据挖掘工具、个人知识管理平台等,同时推进标准化的建设,以实现不同平台间资源的共享。

人才培养与储备中心:e-Science 环境下的文献情报服务对服务提供者的要求将更高,支持中心应该承担起人才培养的责任,培养为科学研究提供服务所需的学科型、通用型、复合型的人才。通过正规教育、在职教育、联机教育等方式实现人才的培养,同时要注重对研究人员信息素质的培养,提高其自我服务的能力。

协调调度中心：**e-Science** 环境下的文献情报服务将更多的以一种虚拟的方式存在，服务团队的组成也将是虚拟的，一项服务的提供可能会涉及若干机构中的人员。因此，需要对这种组织方式进行有效地协调和管理，支持中心通过人力、技术等方面的支持，可以更容易、更方便地执行这种协调调度的功能，成为联合、协同服务中的组织者和协调者。

理论支持、研究中心：**e-Science** 环境下，一方面，需要对已经完成或正在进行的工作进行总结，需要从实践经验到理论的升华，也需要理论指导实践。另一方面，伴随着服务环境的变化和用户需求的变化，文献情报服务也在不断向前发展，因此，需要对这种发展趋势和走向进行研究，对未来的服务可能用到的理论、技术进行研究，如对标准规范的研究、先进技术的研究、发展趋势的研究、政策法规的研究等。这些研究工作在很多时候并不能对科学研究产生实际的作用，不能直接解决科学研究中的问题，但却是文献情报服务可持续发展必不可少的。鉴于其对科研作用的间接性、投入大、周期长，因此需要实力相对雄厚的综合性文献情报机构承担。

13.2.3 服务手段的改进——实现零障碍服务

实际上，在网格技术出现之后，科技界开始憧憬这样一种科学研究环境：科学家能够随时随地获得所需要的信息和数据；随时与大洋彼岸的同事讨论问题或一起做实验，就像身边的同事一样；足不出户地使用分布在全球的科学仪器设备和计算资源，而使用所有的计算、数据、信息等资源就像使用电力一样简单，只需要插上插头就可以了，这就是对 **e-Science** 环境的理想描述。

在 **e-Science** 出现之前，用户与资源、服务之间相隔的是一堵镂空的墙壁，墙壁上有无数的孔洞，不同的资源和服务就从这些孔洞流向用户。用户在使用之前首先面临选择哪个孔洞的困惑，当用户千辛万苦获得了结果之后，尤其是资源之后，还需要判断资源的质量如何、内容是否符合需要。**e-Science** 的出现使这堵墙壁发生了变化，无数的孔洞融合成为一个，用户不再需要选择，所需要的资源和服务将从这个唯一的孔洞源源不断的流出，用户也不再需要对结果进行甄别，因为结果已经根据用户的个性化特征进行了过滤和处理。这就是 **e-Science** 环境下的科学研究，也是 **e-Science** 环境下科学研究对文献情报服务的要求。在 **e-Science** 环境下，融入科研过程之中的文献情报服务，需要从下几方面着手：

1. 一站式服务——满足科研用户对资源和服务的无缝存取

零障碍服务的直接体现是科研人员对资源和服务的无缝存取，为此，资源集成是基础，最终要实现服务的集成。

无疑，**e-Science** 环境下的文献情报服务将主要通过网络来提供。就目前的现状来说，用户在通过网络享用由不同系统提供的服务时，遇到的一个障碍是注册或缴

费问题。一些系统通过注册收集用户信息，希望能够借此了解用户的特点和需求，一些付费系统更是通过注册控制用户的合法使用，但是这种机制却给用户带来了使用成本上的增加。从经济学的角度来说，当用户感觉到其享用服务的成本（包括经济上的、时间上的、技术上的）超过了享用服务所带来的收益时，或者获益不大时，他（她）就可能放弃这项服务。用户对服务的兴趣是不同的，有许多服务可能不会经常使用，甚至只使用一次，因此，当数字科研环境到来时，这种各自为政的机制将极大地限制用户对服务的享用，成为文献情报服务融入科研过程的阻碍因素。

在 e-Science 环境下，可以建立一种统一的注册和管理机制来缓解这种矛盾，开发注册管理系统，使用户可以“一码走天下”，用户在这个系统注册并预先支付一定的费用后，就可以获得一个用户身份，利用这个用户身份，用户可以享用在这个注册系统中登记的所有服务，而费用的结算、支付由系统根据用户的使用情况自动完成。这就为用户屏蔽了使用不同机构提供的服务时的需求差异，同时也兼顾了不同机构的利益，它们可以从系统中获取所需要的用户信息和服务收益。

2. 服务集成

在目前的文献情报服务中，服务集成也是一种服务手段，但目前集成的主要是资源，是在文献情报服务内部的集成，主要解决的是跨库检索的问题。e-Science 环境下的集成并不仅仅是上述资源的集成，而是文献情报服务与用户的科研过程的集成，使之成为科研活动的一个有机组成部分，成为用户知识管理的一部分，像仪器设备、计算资源一样，成为数字科研环境中一个不可或缺的结点，让科研人员实现对资源和服务的无缝存取，在需要的时候能随时随地获取所需的任意资源和服务。

集成的目标是实现文献情报服务与科研活动的融合，要让科研人员既感觉到文献情报服务就在身边，同时又感觉不到文献情报服务就在身边。前者说明文献情报工作的重要性，一旦失去了它，科研工作就无法顺利进行下去，就像停电会使我们的生活陷入混乱一样；而后者则说明文献情报工作已成为科研工作的一部分，已无法将文献情报工作从科研过程中剥离和区分出来，科研人员在从事科研活动的时候已经不再区分哪些是文献情报工作，哪些是科研活动，而是作为一个整体运行着。

3. 协同服务——文献情报服务的生存方式

e-Science 环境下的科研活动将突破单位界限，是一个虚拟的合作团体，这就要求为其提供服务的文献情报部门联合起来，协同服务。e-Science 环境下的协同服务与传统的联合服务的不同之处在于：协同的范围将更加广泛；协同的内容不局限于资源的共建共享；协同的形式将是虚拟的，更加灵活；协同实现的平台是网格。

联合是实现 e-Science 环境下协同合作的主要形式，e-Science 环境下的联合其范围将是广泛的，联合的结果将是双赢的。无论是松散型的合作还是紧密型的联盟，无论是基于资源的还是基于问题、任务的联合，其目的都是为了更好的提供服务，

满足用户需求。可以从以下四个角度来实现联合：

- 纵向合作：与处于不同层面的文献情报机构合作。如与上、下级文献情报机构的合作。
- 横向联合：地区内或不同地区的文献情报机构间的合作，与全球间的科研合作相配合，发挥各自优势，尤其是地域优势，提供服务。
- 行业内联合：联合本行业、本领域的文献情报机构，做好学科信息服务。
- 跨行业联合：与其他行业、领域的文献情报机构联合，与其他性质的机构如公司、企业联合，例如与出版发行行业的合作，与数据库提供商的合作，与 IT 业的合作，与媒体的合作等。

这些联合形式常常是交叉存在的，通过联合可以汇聚力量，扩展资源储备，扩大学科覆盖和地域覆盖，增强核心竞争力和服务能力。

4. 个性化服务

所谓个性化，就是要针对不同用户群体的需求特点，为其提供专门的服务。例如，不同的科研人员向文献情报服务系统提出了相同的问题，但系统反馈给他们的解答结果并不一定完全相同，这种不同体现在所包含的内容可能不同，组织的方式可能不同，表现的形式可能不同，而产生这种不同的根源在于用户个人的兴趣、爱好、研究方向等方面的个性化差异。

个性化推荐、个性化定制、个人网站……个性化实现的形式有许多种，但从用户的角度看，它是文献情报工作所提供的多种形式的服务在用户层面上的统一，实质上是科研人员个体的 Mylibrary 或科研团队群体的 Ourlibrary 与科研平台的集成。因此，e-Science 环境下，个性化服务应该在两个层面上实现：对科研人员个体的个性化服务和对科研团队群体的个性化服务。

首先，这种 Mylibrary 或 Ourlibrary 是文献情报工作者帮助科研人员建立的，是文献情报工作的知识管理系统与用户个体或团队的知识管理系统的融合，用户可以通过这个系统获得它所需的任何资源和服务。

其次，Mylibrary 与 Ourlibrary 之间并非简单的包含关系，二者之间有很大的交叉和互补。因为一个个体可能同时属于多个虚拟团队，其 Mylibrary 中会包含与这些团队相关的内容，从一个个体角度来说，他所研究的内容彼此应具有强相关性，这种相关性意味着其 Mylibrary 中的内容彼此具有强相关性，可以互相补充。例如 A 团队中的个体 a 的 Mylibrary 中的与其他团队相关的内容或在参与其他团队的研究过程中建立的内容可以作为 A 的内容和服务的补充。另一方面，一个团队的成员一般不止一个人，不同个体的 Mylibrary 中有重复，也有不同之处，它们之间是一种补充，而团队的 Ourlibrary 必然会有许多其成员的 Mylibrary 中不包含的内容，这些内容对个体来说也是一种补充。

在 e-Science 环境下，个性化服务的目标是为用户建立一个融入用户科研平台的

Mylibrary 或 Ourlibrary, 集成用户所需的所有资源和服务。对于每一个用户或科研团队而言, 他所面对的由不同机构提供的文献情报服务有一个统一的入口, 就像本节前面所描述的那样, 而不是各个分割的系统各自提供着自身的服务。

13.2.4 人员构成的虚拟化——虚拟动态团队的人员构成机制

e-Science 环境下的科学研究是基于虚拟组织的科学研究, 科研团队的组成将是跨学科、跨机构、跨地域的、虚拟的、动态的。科研团队这种动态、灵活的组织形式要求为之服务的文献情报工作的组织也应该是快速、灵活的。在人员组织上, 虚拟团队将是实现 e-Science 环境下文献情报服务的最佳方式。所谓虚拟团队是指一群人为了实现共同的目标而跨越时间、空间和组织边界通过技术进行的互动组织。它实质上是由任务驱动的成员组成的临时团队, 团队成员在时间和空间上具有各自的独立性; 团队内部通常采用面对面交流与基于信息技术的交流相结合的模式进行沟通; 虚拟团队由一个核心层与一个支持层组成, 核心层是虚拟团队运转的轴心, 承担着虚拟团队的组织工作, 维护它的正常运转。而支持层则由若干完成任务所需的专家和工作人员组成。

1. e-Science环境下虚拟团队的特点表现为

- 开放性: 团队可以与其他团队(包括科研团队和服务团队)有机融合;
- 动态性: 团队成员, 尤其是支持层的成员, 在加入或离开团队时更加灵活和自由。因此, 团队的组成具有动态性, 这有利于充分发挥个体的专业知识和时间灵活性, 实现整体大于部分之和的效应。
- 团队组成人员的结构具有复合性和层次性: 团队中既有具有极高专业知识水平的专家, 也有具备相关领域知识的通才, 团队成员利用自身知识相互协作, 解决项目中的问题。

2. e-Science环境下, 文献情报工作可以通过两种形式的虚拟团队为科研提供服务

- 按学科/专题组建的长期团队: 它们的生存期并不因某一项具体任务的结束而结束, 而是在一个较长的时间内存在。在没有具体任务时, 团队成员在全球各地持续监视着学科/专题领域的活动, 跟踪学科的发展方向, 把握学科的前进脉络。在任务来临时, 整个团队或团队中的一部分可以迅速进入角色, 提供服务, 它们是临时团队组建的基础。
- 面向任务的虚拟团队: 这种团队是为完成某个具体任务而临时组建的, 其生命期以任务的完成为界限。根据需要在任务完成的不同阶段吸收不同方面的人员加入而组成一个动态的临时团队。

组建虚拟团队的难点在于如何使团队成员之间的磨合时间最小化, 如何消除成

员之间的文化差异等各种差异，增强团队成员间的信任，使他们在最短的时间内达到最大的默契。

13.2.5 e-Science环境下文献情报服务的保障机制

1. 技术基础分析

e-Science 是建立在以网络为基础和核心的一系列相关技术之上的，要做好这一环境下的文献情报服务，同样需要这些技术的支持。

根据各种技术在 e-Science 发展中的不同位置与不同作用，可以将 e-Science 关键技术分为四类：基于网络的资源和服务共享技术；科研数据的生产、采集、掌管、保存和分析的技术；科研对象的建模和仿真技术；动态的、分布式的、虚拟的研究团队的组建和协同技术。这类技术包括：VO 技术、学术交流技术、问题求解环境、协同诊断和协同决策技术等。

2. 标准化机制

标准化是实现异构系统互操作的另一个技术手段，它是“为了在一定范围内获得最佳秩序，对现实问题或潜在问题制定共同使用和重复使用的条款的活动”。

e-Science 环境下的资源体系是异常庞大和分散的，要实现这些资源的共享，标准化问题至关重要。另外，e-Science 环境下的文献情报服务是由分布在不同机构、不同国家、不同地理位置的文献情报工作者共同完成的，为了保证这些文献情报工作者之间的协作，为了使大家的工作成果能够被共享、利用，更加需要一定的标准规范来约束人们的行为。

标准化建设应遵从科学性、系统性、可扩展性、兼容性、开放性的原则。e-Science 环境下文献情报服务的标准化建设其内容涉及方方面面，如数字资源的描述方法（标识语言）、组织体系（分类方法、开放系统互连标准）、存储格式（文本、图像、音频、视频、多媒体、超媒体等不同类型数据的存储格式），各种格式之间的转换标准、压缩标准、数据交换标准（接口的标准化），资源的数字化加工标准，对知识组织的各种标准、对其他种类的资源组织标准（如软件资源）等。

3. 人才保障机制

合理的人才结构是 e-Science 环境下文献情报服务得以有效运行的基本保障。知识服务的提供对服务者的要求更高，需要专业知识、技术知识、文献情报知识等多方面知识的支持。而要一个人同时具备这么多种知识是困难的，或许终其一生也只有少数人才能达到这一境界，因此，e-Science 环境下的文献情报服务，需要组合人才与人才组合的合力作用来实现。

所谓组合人才，就是具有提供文献情报服务所需的各个领域知识的人才。在人

才的培养上要向复合型方面发展，要具有一定的学科背景，努力使其知识结构在深度和广度上向着文献情报工作所需的标准靠拢，学科馆员是人才培养的发展方向。

所谓人才组合，就是在完成任务的团队中要吸纳各种类型的人才，而不应该局限于文献情报领域。

4. 文献情报服务的评价机制

评价机制的建立是为了及时的获得反馈，发现问题，以便更好的改善服务。按照克里斯丁·格朗鲁斯的观点：服务一般是以无形的方式，在顾客与服务职员、有形资源商品或服务系统之间发生的、可以解决顾客问题的一种或一系列行为，因此，对服务的评价集中在服务内容、服务质量和服务效果三个方面。

评价的方法应采用主观与客观相结合的方法。主观方法是自身对服务的评价，在 e-Science 环境下，这种评价可以通过提供服务的数量（如建立的知识发现工具的数量，对多少数据进行了挖掘，获得了多少知识元和知识等）和用户的使用量（如点击率、浏览率、下载率）两个方面来实现；客观方法是他人对服务的评价，实质是用户的评价，是用户对服务的反馈意见。

对于服务行业来说，用户满意度始终是评价服务效果的最直接、最核心、最有效的指标。文献情报服务也不例外。用户满意度是获得用户反馈的最直接的方法。e-Science 环境下文献情报服务的客观评价将以用户满意度为核心，而用户满意度的具体体现应该包括用户现实需求的满足率、用户潜在需求的激活率和用户问题的解决率。用户满意度的实现方法多采用问卷调查的方法，而实现问卷调查的关键是如何使问卷的设计和填写能够真实地反映用户的需求。

第 14 章 e-Science环境下的数字图书馆

14.1 e-Science环境下数字图书馆范式的演变

随着 e-Science 这一全新的科研协作模式所带来的技术、标准等给数字图书馆各个核心要素所造成的变革式影响，由此引发的用户需求以及整个用户信息环境的波动将直接导致数字图书馆范式发生演变，并促使数字图书馆在功能定位上作出调整。尽管世界上有许多国家开展数字图书馆建设，但由于信息技术和相关标准的不断发展，到目前为止，数字图书馆并没有一个通用的或者是固定的体系结构，相反，其体系结构始终处于动态变化之中。作为数字图书馆范式的重要标志，体系结构的变化也不断地召示着数字图书馆范式的演变、发展。

从出现伊始的初步发展到 e-Science 环境下的快速进步，数字图书馆服务能力不断得到提升，功能进一步完善，范式也已经走过了两次演变并正在经历着第三次范式演变的成熟与完善（详见图 14-1）。以 NCSTRL（Networked Computer Science Technical Research Library）、ELINOR（Electronic Library Information Online Retrieval）等早期较有代表性的数字图书馆项目为例，这些项目的体系框架表明了数字图书馆第一代范式聚焦于数字化资源；到了第二代，数字图书馆范式转向强调服务集成化，以 NSDL、CDL（California Digital Library）以及 CSDL 各自所建立的体系框架充分说明了这一点；在 e-Science 环境下，以网格技术作为支撑的 DILIGENT 项目将自身的目标定位为，为用户创建一种可用于构建个人虚拟数字图书馆的基础设施。利用这一项目的研究成果，今后的数字图书馆范式、构建模式将发生极大改变。事实上，DILIGENT 的到来已表明，e-Science 环境下，数字图书馆新范式——虚拟数字图书馆正日趋成熟。

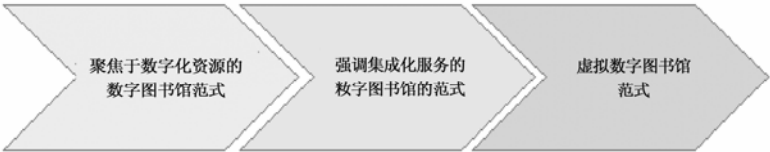


图 14-1 数字图书馆范式演变的轨迹

14.1.1 聚焦于数字化资源的数字图书馆范式

1. ELINOR及其体系框架

ELINOR 项目是英国德蒙特福德大学（De Monfort University）、英国图书馆和

IBM 英国公司合作进行的试验性项目，始建于 1992 年，重要目标之一是构建一个数字图书馆原型系统，使德蒙特福德大学大学师生能通过校园网络利用个人电脑直接存取经常被使用的图书、期刊、课程资料的全文以及多媒体学习资料等。

图 14-2 表明，ELINOR 数字图书馆中服务器功能的实现是通过搜索引擎、文献输入系统、使用统计数字收集与管理子系统、打印控制子系统以及工具包等协同完成的。在这一体系框架中，输入系统主要用于处理两类资源——印刷文献与电子文献，为数字图书馆开展服务提供资源支撑。

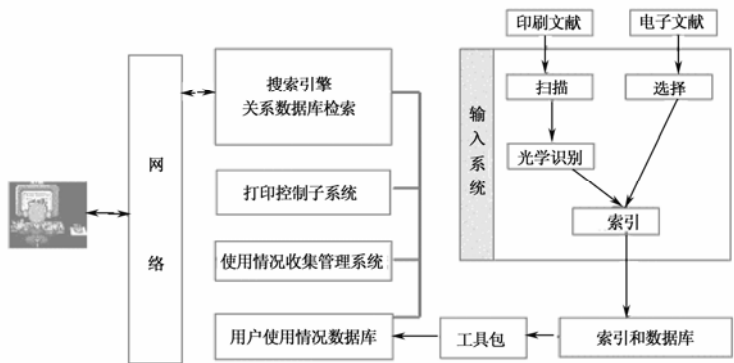


图 14-2 ELINOR 的体系框架

2. NCSTRL 及其体系框架

NCSTRL 是一个广泛应用的数字图书馆项目。它的建设目标是在 Web 上建立一个计算机科学研究资源的全球数字图书馆。它始建于 1994 年，到 2001 年 9 月它一直由康奈尔大学 (Cornell University) 的数字图书馆组织负责执行和维护。截止于 2001 年，NCSTRL 已经连接了分散在美国、欧洲和亚洲的 100 多家大学和研究机构，收集了多家大学以及研究实验室生产的有关计算机科学方面的大量科技报告。

通过分析图 14-3 可知，浏览器、命名服务、用户界面、资源库、查询路由和藏品服务是 NCSTRL 体系结构的基本组成要素。事实上，在 NCSTRL 的早期项目中，用户的查询式被直接提交给所有的索引服务器，但随着参与机构的增多，检索时会因为其中的一个元数据库的查询失败或响应较慢，从而造成整个检索过程的失败或长时间的响应，为了解决这一问题，NCSTRL 引入了查询路由器的概念。在整个建设过程中，资源库始终是它的重点。

3. 特征分析

以 ELINOR、NCSTRL 为代表的数字图书馆项目及其体系框架表明，在这一发展时期，数字图书馆的建设模式是，在某一特定文献资源数字化的基础上建立数字信息资源系统，而且这些系统往往被作为独立系统嵌入到传统图书馆系统或上层机

构信息系统中，并将跨时空检索和传递特定数字化资源作为建设的主要任务。资源库（用于存储数字化资源）、资源管理与检索、用户界面是这一时期数字图书馆范式的全部组成，而且数字化资源（各类资源库）是这一时期数字图书建设的首要出发点，更是这一代数字图书馆范式的突出特征。

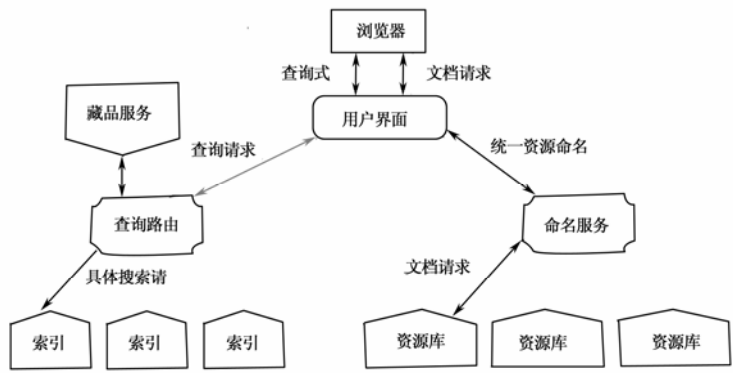


图 14-3 NCSTRL 的体系框架

14.1.2 强调集成化服务的数字图书馆范式

1. NSDL及其体系框架

1998 年，美国国家科学基金会正式启动了 NSDL 项目。按此计划，NSDL 将发展一个联机环境，通过因特网向各层次的学生和教师提供高质量科学、数学、工程与技术教育资源。NSDL 一期工程已于 2002，年底为公众提供服务，它是目前规模最大的数字图书馆项目。由于 NSDL 的内容和用户的多样性，为让各种用户共享不同的信息，NSDL 设计的体系框架以共享元数据为出发点（见图 14-4），并利用元数据开发核心服务（如搜索和发现）。

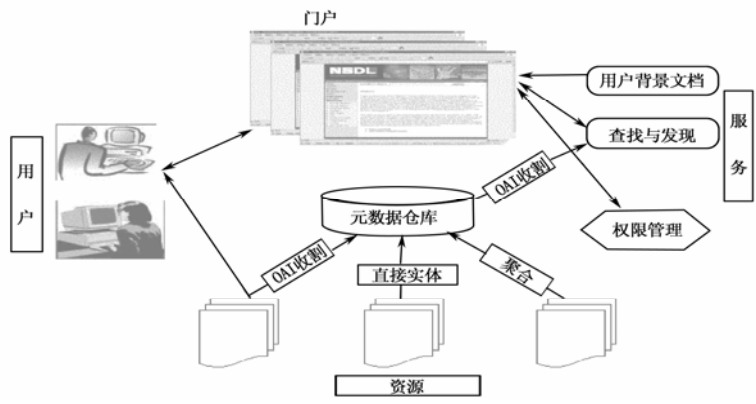


图 14-4 NSDL 的体系框架

分析图 14-4 可知, NSDL 采取高度集中与高度分散的原则, 即, 利用一个门户网站, 可以为学生、教师、科研人员等不同类型的用户服务; 同时, 它强调基于广泛资源的服务, 力求将可以提供服务的资源扩展到文献资源以外的资源; 而且, 它将自身的建设基础直接架构在新技术、新平台之上, 保证服务的可靠提供; 此外, 为了保证提供给用户的资源在质量上的可靠性, 它会提供与资源最为相关的链接。

在开展服务的过程中, NSDL 希望自身可以成为一个科学教育创新中心。为实现这一目标, 它将通过五个方面进行努力。即, 营造合作环境, 建立社区中心, 以便开展知识和资源的共享; 收集用户有价值的反馈, 不断提升服务质量; 与教育界同仁开展合作; 对教育环境和政策施加影响; 在资源的建设过程中, NSDL 的用户既是资源的接受者, 同时也是 NSDL 的资源供给者。

为保证自身可以在核心集成、资源建设、服务提供与专题研究等方面发挥服务优势, NSDL 将元数据登记系统、数据库仓库、检索与发现服务系统、存取管理服务系统、核心存取系统、用户界面和门户、存档服务系统、用户帮助系统、参考咨询系统九个方面有机纳入自身的体系构成中。

2. CDL及其体系框架

CDL 是第一家研究型图书馆, 也是加州大学的数字图书馆, 始建于 1997 年, 由 11 家学校合作组建, 它采用商业运作手段, 实施平等管理。同时, 它更是一个商业模式, 有多名员工一起合作提供服务。

CDL 采用 InterLib 集成服务, 系统具有智能检索电子文献资源、查询多媒体资源的服务功能。2000 年, CDL 在阐述自身的功能时指出, CDL 是一个学术信息管理的集成数字系统。它直接服务于校园内的教职工和学生, 随后逐步将服务对象扩展到普通市民。CDL 的体系框架表明 (见图 14-5), 依托高质量的电子知识资源, CDL 通过网络工具将它所能够提供的各类分布式服务通过网络传递到用户桌面上。

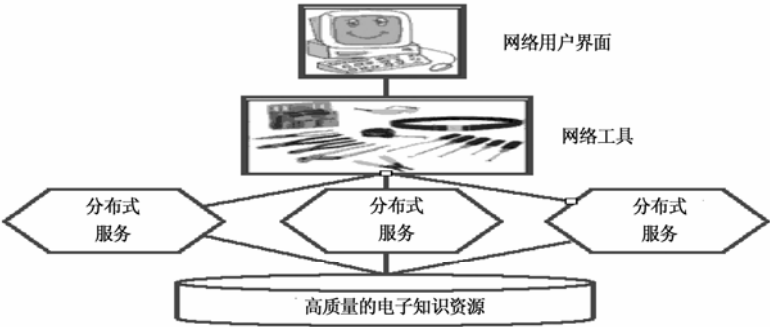


图 14-5 CDL 的体系结构

3. CSDL及其体系框架

CSDL 的目标是建立和维护中国科学院全院网络共享的科技信息保障环境, 提供全院“一体化”和“一站式”的科技信息服务。为支撑各类服务、保证功能的正常发挥, CSDL 采用了三层体系模式 (详见图 14-6), 其体系框架由资源层、功能服务层以及服务集成调度层三个模块组成。利用这种面向集成化服务的体系框架, CSDL 能够方便地、快速地将多种服务提供给用户使用。

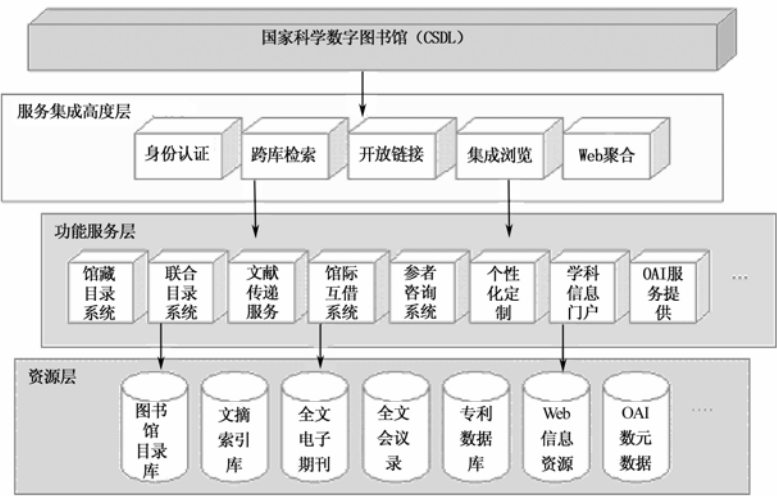


图 14-6 CSDL 的体系框架

目前, CSDL 的一期建设已得到全面实施。在资源建设方面, CSDL 提供中外文的电子期刊、会议录、学位论文、专利、科学引文索引和学科信息门户等类型的文献数据库 128 个, 涉及多个学科领域等, 用户利用其跨库检索引擎可直接查询上百个全文、文摘和馆藏目录数据库, 并获取电子版全文; 利用其文献传递系统与国家科技图书文献中心网络服务系统, 用户可在 48 小时内获得 18000 种西文期刊的全文传递服务; CSDL 的随易通服务, 使用户在任何地点上网, 均可查询所在单位购买开通的数据库。利用它提供的 My Library 服务, 用户可以在学科信息、参考书架、搜索引擎、全文数据库、文摘数据库之中进行服务定制。其虚拟参考咨询服务水平也正在不断改进, 提供实时、延时和专家咨询等服务和二级咨询机制保障。学科信息门户、特色门户以及热门门户的建设也已初具规模。

4. 特征分析

以 NSDL、CDL、CSDL 为代表的数字图书馆项目及其体系框架表明, 在这一发展时期, 数字图书馆的建设模式是, 以前期的资源建设为基础, 致力于支持分布的数字信息系统间的互操作, 为互操作系统之间的无缝交换和共享信息资源提供服务,

在此基础上构造一种逻辑的集成信息服务机制。在此时期，数字图书馆范式的组成要素是各种分布式资源、提供集成的工具、集成化服务。其中，集成化服务是第二代数字图书馆范式的强调点。

14.1.3 虚拟数字图书馆——e-Science时代的数字图书馆范式

1. 虚拟数字图书馆出现的背景

在 WWW、因特网技术出现之前，科研主体开展科学研究主要依靠传统的科学信息和出版物作为科研的支撑来开展科学研究，当借助于 WWW、因特网技术以及适合于在线信息发布的文档格式标准而建立起来的在线数字图书馆出现时，科研主体便可以依靠它实现全球化的信息高效获取。当人类历史发展到今天可以利用 SOA（Service-oriented architecture）、网格基础架构、先进的数字图书馆服务以及其他各类资源服务时，科研主体就已进入了一个被称为 e-Science 的环境。这是一个与虚拟信息与知识环境、电子出版、虚拟数字图书馆服务提供密切结合的环境，在此环境下，科研主体将通过获取具有明显个性化特征的下一代知识产品与知识服务来满足科研过程的需求（图 14-7）。

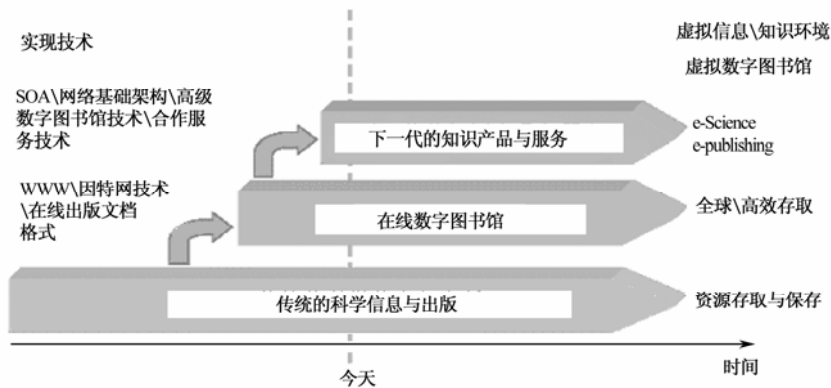


图 14-7 走向 e-Science 的数字图书馆

IPSI（Integrated Publication and Information Systems Institute）研究者 Matthias Hemmje 指出，在传统的信息与出版模式中（如图 14-8 所示），科学家在科研的过程中向数字图书馆提出信息需求，而数字图书馆作为服务提供商开始调动自身的资源去满足科学家的需求。借助于数字图书馆，科学家个体与研发团队进行交流。与此同时，研发团队也存在着各种各样的信息需求，也需要向数字图书馆寻求服务。在满足单个用户需求和群体用户需求的过程中，由多个数字图书馆通过一定规划集合而成的数字图书馆联盟也开始出现了。

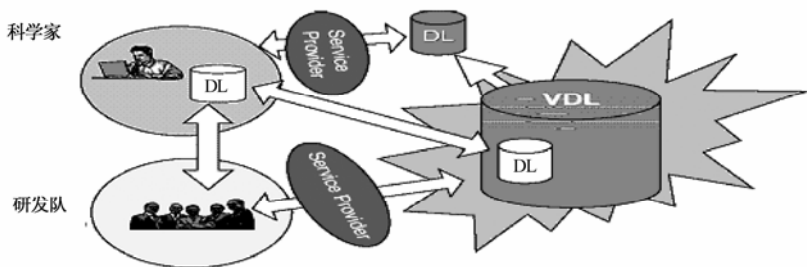


图 14-8 传统的信息及其发布模式

在 e-Science 环境下，科学家个体依然需要与数字图书馆进行信息上的交流，而研究团队也需要向数字图书馆获得信息需求上的满足。所不同的是，在 e-Science 背景下，数字图书馆已不再单纯是信息服务的主体，更是身兼信息服务的客体。这就意味着，以网格为基础设施的数字图书馆正逐渐向虚拟数字图书馆演化。在此过程中，研究所、大学等单位将成为虚拟数字图书馆的内容提供商和服务提供商，同时，虚拟数字图书馆利用这些内容提供商和服务提供商的资源与服务，进一步完善自己，最终使得自身能够以更为完美的服务提供商身份出现在科学家个体和研发团队面前，给他们提供更完美的服务（参见图 14-9）。

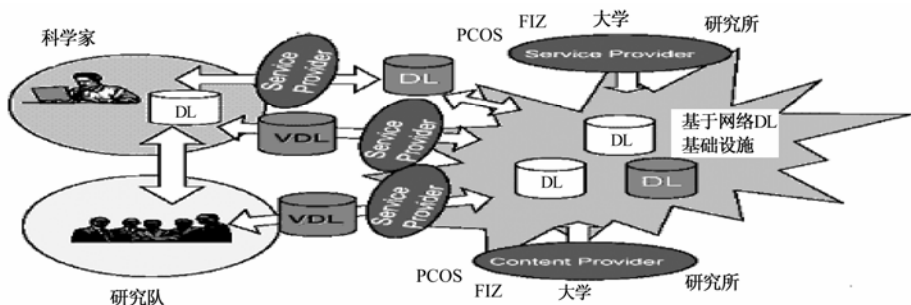


图 14-9 e-Science 环境下的信息及其发布模式

2. 早期的虚拟数字图书馆建设

事实上，早在 e-Science 出现之前，有关虚拟数字图书馆的问题就已成为数字图书馆界的讨论点。2000 年，美韩数字图书馆双边研讨会就虚拟数字图书馆的原型架构展开了讨论。经过激烈的争论，与会者一致认为，建立虚拟数字图书馆，其目的乃是为涉及教育或者文化遗产的数字图书馆提供支持。一般而言，虚拟数字图书馆是一种数字对象的动态集，而且这些数字对象可以让人们通过当前的或者是动态生成的界面加以获取。除了理论上的探讨，欧洲在 1998 年启动了虚拟数字图书馆建设项目。

欧洲虚拟数字图书馆建设项目 VILIB (Virtual digital LIBrary) 可以说是最早的虚拟数字图书馆项目。该项目由欧共体支助, 持续时间为 1998 年 1 月到 1999 年 12 月。项目计划开发出一个数字图书馆, 其规模横跨整个欧洲, 研究成果支持使用者对连接到互联网并且支持 Z39.50 图书馆协议的各个欧洲图书馆的目录名称和全文文档进行“跨语言”检索, 用户可以使用自然语言进行检索。

在检索的过程中, 用户可以根据自身的需要选择图书馆目录, 也可以选择他所喜欢的语言提交检索词, 更可以利用系统提供的自动翻译功能将这些检索词翻译为某种语言后再进行检索, 实现跨语言检索。在检索过程中, 系统将从检索式中提取关键词, 并使用与检索式相同的语种来表达关键词进而实施检索。与此同时, 如果用户使用了跨语言检索, 那么, 系统也会根据用户所选择的目标语种对关键词进行翻译, 之后再进行检索。这样一来, 检索结果使用的语种将与用户所选择的目标语种一致。比如说, 用户输入中文检索词并希望得到英文资料, 通过利用该检索系统, 用户的这一目标将可以轻松实现。到目前为止, VILIB 项目提供的、可以实现跨语言检索的语种有德语、法语、英语和西班牙语四种。在搜索图书馆目录时, 用户将使用 LIBRARIES 项目所开发出来的系统进行搜索。VILIB 系统使用 WWW 的用户界面。用户通过 Web 页提交检索式, 并经由采用 html 语言编写的 Web 页获得检索结果。利用这种用户界面, 用户不仅可以获得相关文献资料的标题、全文, 还可以获取图像、声频和视频等资源。

尽管 VILIB 较早就投入研究与建设, 但是, 却因技术支撑的不完善使得自身并不具备明显的虚拟数字图书馆特色。

3. e-Science环境下虚拟数字图书馆的原型项目——DILIGENT

(1) DILIGENT 概述

DILIGENT 项目由欧盟资助 630 万欧元于 2004 年 9 月开始启动, 为期 3 年, 项目的目标是, 创建一个先进的试验床, 使动态虚拟的 e-Science, 组织成员能够存取共享的知识, 并以安全、协调、动态和低成本、高效益的方式进行合作。来自 ISTI-CNR (the Institute of Information Science and Technologies National Research Council) 的 Donatella Castelli 指出, 启动 DILIGENT 项目是为了创建一个由网格技术支持的先进数字图书馆基础架构。该架构允许来自动态虚拟组织的成员在共享计算资源、存储资源、多媒体资源、多类型内容资源以及多类型应用资源的基础上建立一个临时的虚拟数字图书馆。DILIGENT 试验床建立在网格技术与数字图书馆技术加以集成的基础之上。这种集成将为部署下一代 e-Science 知识基础设施打下坚实的根基。

(2) DILIGENT 的建设特色

① 面向 EGEE 的建设需求。EGEE (The Enabling Grids for E-science) 由欧盟委员会资助, 汇集了超过 27 个国家的专家, EGEE 项目的目标是基于网格技术取得的最新进展, 开发一个欧洲的服务网格基础设施, 可供科学家全天 24 小时使用。该

网格获得欧盟 3000 万欧元的资助，是同类网格项目中最大的一个。该项目将主要集中于三个方面：建立一个一致、健壮和安全的网格，可以吸引其他计算资源的加入；不断地改进和维护中间件以提供给用户可靠的服务；吸引科研和工业领域的新用户，并确保他们获得高标准的培训和支持。作为欧洲开展 e-Science 的基础设施之一，DILIGENT 将自身的建设置身于 EGEE 的建设环境中并以 EGEE 提供的服务保证功能的正常发挥（图 14-10）。

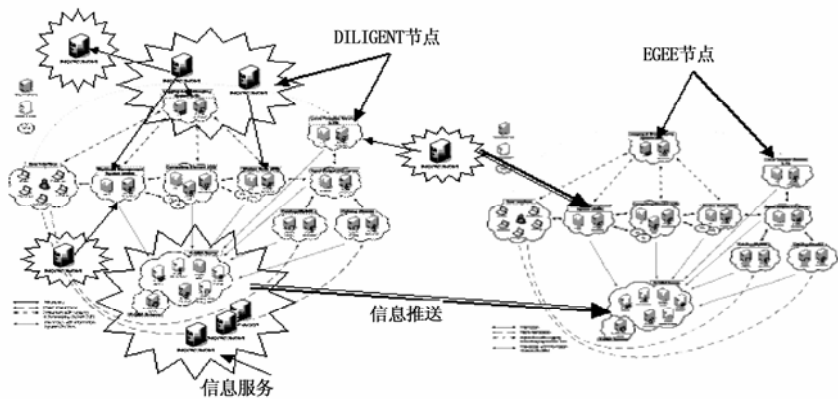


图 14-10 面向 EGEE 的 DILIGENT

② 遵循 OGSA 协议。OGSA，主要是将 Web Services、数据库存取、J2EE 等技术规范纳入网格计算，使得网格服务统一以 Services 的方式对外界提供。在 DILIGENT 构建的系统架构中，与 OGSA 协议相兼容的中间件是其资源层的重要组成部分。通过这种方式，DILIGENT 确保自身提供的服务均遵循 OGSA 协议，让第三方可以将 DILIGENT 提供的各类服务以统一服务的方式进行集成并提供给最终用户使用。

③ 瞄准 e-Science 环境下数字图书馆的发展趋势。在未来，无论是个人学习，还是从事科学研究，均需要数字图书馆的支持。面对数字图书馆的发展趋势，来自 ISTI-CNR 的学者指出，数字图书馆未来的发展走向集中表现为三点：支持内容与服务资源的共享；数字图书馆系统具有保证数字图书馆功能正常发挥所需要的一切关键服务与管理特征；数字图书馆作为通用共享基础设施的重要组件。DILIGENT 的建设将以此为重要导向，力求站在未来数字图书馆的发展前沿。

事实上，除了 DILIGENT 项目，欧洲已经有多家组织与研究机构开始致力于网格技术与数字图书馆技术结合的研究。例如，DELOS（a Network of Excellence in Digital Libraries）网络组织实施了一个联合的计划，其目的在于整合与协调欧洲地区从事数字图书馆领域的一些主要团队正在进行的研究工作。目前，这些团队的目标是发展 e-Science 环境下的数字图书馆技术。这种技术将支撑任何用户在任何地方的任何时刻以一种友好的、多模式可选的、高效的方式去存取人类知识。不过，在成

熟度方面，DILIGENT 构造的系统框架及支撑技术均处领先地位。

总之，DILIGENT 以 EGEE 为支撑，底层采用与 OGSA 兼容的构件组建，建设理念与服务机制符合 e-Science 环境下数字图书馆的运作模式。

(3) DILIGENT 技术解决方案

图 14-11 表明，DILIGENT 以 EGEE 项目的建设成果为依托，其技术解决方案主要涉及七个方面。首先是虚拟数字图书馆的创建。即是说，在 e-Science 环境下，用户在寻求可以满足自身需求的资源前可以根据个人喜好、研究背景自行构建个人虚拟数字图书馆；其次是个性化服务、流程设计与验证、资源搜索、内容与元数据的管理、注解、内容来源的描述方面的服务。在以往，上述七方面有可能只是由作为服务提供方的数字图书馆单独完成。而在 e-Science 的环境下，随着支撑创建个人虚拟数字图书馆的各项技术逐渐走向成熟，上述七项操作更有可能是由用户特别是科研主体独立来完成。

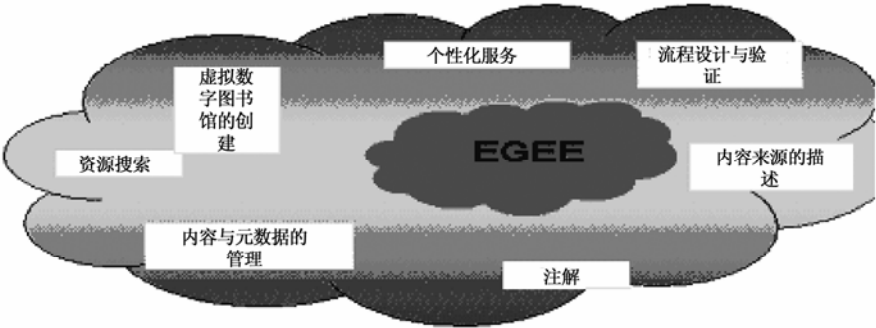


图 14-11 DILIGENT 的技术解决方案

(4) DILIGENT 的服务流程

如图 14-12 所示，DILIGENT 利用相关组件对资源进行采集后，使用符合资源选择标准的格式描述资源，并分配特定的标识号给已经描述过的每种资源。在此基础上，它还要对资源的权威性做出认证。当用户通过用户界面提出资源需求后，资源选择系统向资源调度程序发出请求，要求将符合用户需求的资源加以汇集，此时，调度程序将资源请求发送给虚拟数字图书馆，接到请求后，虚拟数字图书馆即根据需求内容在自身可调控的资源范围内进行探索，并将可能满足需求的资源标记反馈给调度程序。调度程序经过确认后，将符合需求的资源标记再次发送给虚拟数字图书馆，此时，虚拟数字图书馆再将与该标记相匹配的资源实施融合，融合结果通过用户界面返回给用户。

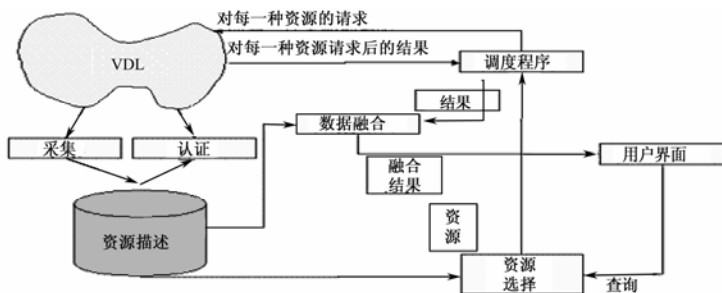


图 14-12 DILIGENT 的服务流程

(5) DILIGENT 功能定位

DILIGENT 的建立，将集成数字图书馆的服务，并以 EGEE 所开发出来的基础架构为基础。随着建设的逐步完善，在 e-Science 环境下，DILIGENT 将可以被无缝地集成到 OGSA 中。在功能上，它将突出显示了数字图书馆与 e-Science 的核心技术——网格技术结合后的特色。DILIGENT 的功能定位包括五个方面：

① 数字图书馆的创建与管理。包括：信息服务、守护者的服务、代理与匹配生成器服务、虚拟组织支持服务、虚拟数字图书馆生成器服务。

② 流程管理。包括：流程设计与认证服务、流程执行与可靠性服务、流程最优化服务。

③ 内容与元数据管理。包括：内容管理服务、元数据管理服务、元数据代理、内容安全、注释管理服务。

④ 索引与检索服务。包括：特征抽取、索引服务、内容来源的描述与选择服务、数据融合服务、搜索服务、个性化服务。

⑤ 终端用户的服务。DILIGENT 创建了一个功能强大的用户界面。一方面，这有助于基础架构的管理与维护；另一方面，这可以方便用户使用，从而使数字图书馆系统的潜能得到全面的开发。为了使得开发出来的系统可以跟绝大多数的门户引擎兼容，DILIGENT 采用各种技术建立起来的图形 portlets 将以各种工业标准为基础。最后，用户与管理者可以通过个性化网络门户提供的、更为直观的操作界面来使用、管理 DILIGENT 平台。

除了提供一种能够让终端用户与系统进行交互的机制，在项目的初期工作计划中，DILIGENT 还将为两个应用领域的主要信息类型提供可视化的基本组件。

4. 虚拟数字图书馆范式的特征

通过分析以第一代、第二代数字图书馆范式特征可得出，e-Science 到来之前，数字图书馆范式组成主要包括四部分。即资源库、命名系统、索引与搜索系统以及用户界面。

(1) 资源库。资源库的功能包括存储和管理各种数字对象，通常是由关系型数据库来管理。应用程序通过 RAP 访问资源库。可实现存储、访问、复制、移动和删

除数字对象等操作。与 Web 上的信息不同, 数字图书馆中数字资源是以数字对象的形式进行封装的, 一个数字对象包括: 一个全球唯一的独立于地址的长期标识符; 数字资料, 存储数字图书馆的资料, 也就是最终用户需要获取的信息内容, 如经 XML 置标后的文本、一本电子图书等; 元数据。

(2) 命名系统。命名系统是针对长期标识符的分配、管理及解析的一个综合系统, 其中以调度系统 (Handle System) 较为有名。它是一个独立的系统, 其职能是负责数字资源的全球唯一的、长期的、独立于地址的命名的分配、管理和解析。在调度系统中, 本地名称空间通过获取一个调度系统的命名授权, 就可以纳入到全球调度名称空间, 这样所有的本地名称在全球调度名称空间中将是唯一的。

(3) 索引与搜索系统。索引的创建可能是由机器的自动扫描、手工录入和干预, 或者是这两者的结合。客户机把查询式提交给索引服务器后, 将返回相匹配的数字对象的定位码。索引服务还提供被索引信息的元数据和查询机制。

(4) 用户界面。用户界面是用户与数字图书馆的接口, 数字图书馆向用户提供的最终服务都是通过用户界面来实现的。一般情况下, 数字图书馆借助通用的 Web 浏览器作为其用户界面工具。此外, 用户界面的内容编排和服务方式问题是很复杂的, 不同的用户需求是不同的, 用户对数字图书馆的简单灵活的需求首先就表现在用户界面上。

如果说, e-Science 出现之前, 数字化资源是一切数字图书馆的生存之本、是数字图书馆体系结构的重要组成部分。那么, 在 e-Science 环境下, 这种局面将被打破。以网格为代表的多种 e-Science 技术, 将使得虚拟数字图书馆能够在极短时间内汇集全球范围内一切可供调用的资源, 包括信息资源、设备资源、仪器资源等, 或者是帮助用户在极短时间内构建个人虚拟数字图书馆并让虚拟数字图书馆自身去收集所需资源。e-Science 环境下, 尽管虚拟数字图书馆将不再以拥有资源作为建设重点, 但是, 只要权限许可, 处处资源都是它可以支配与调度的对象。

提供资源采集、组织与管理的规则与协调机制, 让符合用户特定需要的资源在极短的时间内响应用户需求, 让用户根据自身研究课题、学习需要以及参与 VO 的需要而创建极富个性化的虚拟数字图书馆, 这将成为数字图书馆继以资源库建设为核心之后的一个重要关注点。实际上, e-Science 环境下的个人虚拟数字图书馆在功能上将比当前的个人数字图书馆更强大、操作起来也更有人性化。e-Science 环境下, 虚拟数字图书馆这一新范式将至少包括规则库 (包括用户创建、管理、维护个人虚拟图书馆的规则; 虚拟数字图书馆汇集资源、联合服务的调度规划与响应机制等)、命名系统、索引与搜索服务、数字图书馆生成器与用户界面等组件, 同时面向用户需求、提供数字图书馆生成器服务作为自身的重要标志。

总之, 虚拟数字图书馆最终的目标就是让用户建立起完全满足自身资源需求的 MyLibrary。它与当前数字图书馆提供的 My library 完全不同。这种不同表现在, 它可以从用户自己的计算机出发, 使用户可以使用自我进行定义的标准创建个人虚拟数字图书馆, 把所有的信息动态地集成到一起, 并与 e-Science 环境结合在一起, 最

终使它成为以用户信息活动为基础的数字信息环境。

14.2 e-Science环境下数字图书馆的功能框架

e-Science 环境下，尽管数字图书馆的范式将发生变化，而且功能定位也将有所改动，但是，注重从资源、服务与管理三个角度来体现自身功能依然是它的出发点。e-Science 环境下，数字图书馆依然需要创建与获取资源，但是，这已经不再是它建设中的首要出发点，在用户寻找资源、构建个人虚拟数字图书馆的过程中，数字图书馆将充当资源代理器的角色，为用户提供更多个性化、集成化的服务，向用户传播资源采集、描述、组织、管理的方法、规则与技巧，包括将资源描述与组织的过程进一步向用户开放、提高用户对信息的敏感度等。

虚拟数字图书馆的创建与管理将是 e-Science 环境下数字图书馆在功能上最大创新，也是 e-Science 环境下数字图书馆核心竞争力的表现。作为数字图书馆个性化服务的高级形式，虚拟数字图书馆不仅继续保留、沿用 e-Science 到来之前数字图书馆提供的 MyLibrary 服务，而且会将之继续扩展，在满足用户需求的过程中更注重人性化。

总的来说，e-Science 环境下，数字图书馆的功能框架尽管由三部分组成（如图 14-13 所示），但是，与 e-Science 到来之前的各项功能比较，无论是在内涵上，还是在外延上，数字图书馆的功能都发生了极大的变化。支持用户创建个人虚拟数字图书馆作为 e-Science 环境下数字图书馆的核心功能，它的顺利实现需要包括虚拟数字图书馆生成器（DILIGENT 项目将此作为研究与建设重点）等多项技术的支撑，也需要建立在数字图书馆资源功能、服务功能、管理功能协调发展的基础之上。

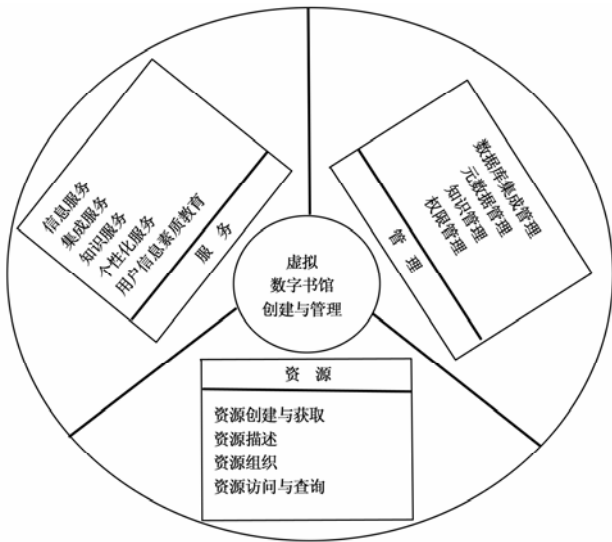


图 14-13 e-Science 环境下数字图书馆功能框架

14.3 演变后的功能特点分析

1. 交互性增强

虚拟数字图书馆将不再只是信息服务的主体，更是身兼信息服务的客体。也就是说，在为用户提供信息的过程中，虚拟数字图书馆也接受用户个人或者是来自 VO 所提供的资源以进一步充实自身的资源。用户在成为资源接受者的同时，也作为资源贡献者加入到数字图书馆的资源建设阵营中。与 e-Science 到来前仅仅是让数字图书馆通过提供参考咨询服务、电子邮件来与用户交流的方式相比，这种方式使得数字图书馆与用户个人、与 VO 之间的交互作用在程度上更加深入。

2. 个性化增强

在 e-Science 环境下，虚拟数字图书馆通过汇集分布于各处的资源，利用网络技术构建集成服务。之后通过提供用户可以自由定制的界面，让用户进行资源与服务的定制。再将经过用户定制后的服务打包（wrapped）嵌入到用户系统中供用户使用。采用网络技术后数字图书馆在个性化服务上功能将更加强大。DILIGENT 项目的建设目的就是建立一个依托网络技术的数字图书馆基础设施。利用这一基础设施，用户将可以方便地构建适合自身某一时期的研究需要的个人虚拟数字图书馆。

3. 动态集成性增强

虚拟数字图书馆的集成服务建设是一个庞大的信息工程，涉及许多方面，只有协同工作，才能保证正常地运转。在 e-Science 环境下，利用其核心支撑技术——网络将分布在不同地理位置的资源通过高速的因特网进行资源集成，从而提供一种集高性能计算、管理及服务于一体的资源服务能力。在分布式的异构环境中，网络技术能够精确定位所需的数据集（data set），并为后续操作提供必要支持。人们利用这些资源就像利用电源一样，不必关注资源来源及系统负载状态。此外，网络技术已发展成为连接和统一各类远程异构资源的重要途径，它可以合理而有效地组织远程资源并组建具有强大服务能力的网络虚拟计算机。利用网络这一技术优势，数字图书馆在建设时，就可以将权限管理、知识库、各项服务进行汇集，并由网络技术实施资源调度，最后以一个统一的访问界面提供给用户访问。利用这一界面，用户可以获取由单个数字图书馆或者多个数字图书馆组成的联盟数字图书馆提供的服务。

4. 知识性增强

与因特网相比，网络具有知识生产特性。即是说，由于因特网并不生产知识，因此，信息知识只能通过因特网以外的其他途径被创造出来后才可以在因特网上进

行传播，供用户查找利用，而网格则可以从用户需求的角度出发，自动生产知识，在此过程中，服务能力强大的高性能计算机发挥了重要作用，它借助特定的程序，把从数据源得到的各种原始数据（raw data）处理为信息和知识。不仅如此，网格还可以根据用户需要，自动查找相关数据的数据源，通过对这些数据源进行综合分析和知识发现，形成新的信息与知识。可见，随着网格技术的不断发展，虚拟数字图书馆的功能在知识性上将得到全面提高，在客户提出请求或查询时，依托网格建立的数字图书馆服务系统将会进行自动处理分析，并把有关的结果——用户需要的相关信息与知识传送到客户登录的节点（node）上，从而使得数字图书馆的知识服务更加完善。

5. 及时性增强

在 e-Science 出现前，数字图书馆的服务只是“游离”于科研活动过程之外。因此，数字图书馆在满足用户需求时，更多的是采取一种“后发制人”的方式，即就是说，在接收到科研人员的信息需求后，数字图书馆才根据需求开始着手响应需求的各种操作。在 e-Science 环境下，由于科研人员、VO 对于信息以及各种资源在响应速度上要求更高，因此，虚拟数字图书馆在提供服务时更需要精确、及时。正是这种需求使得数字图书馆在服务方式上必须改为“先发制人”，即就是说，e-Science 环境下，数字图书馆提供的服务要做到能够根据用户提供的背景文档（包括科研人员的学历、专业与研究方向，正在着手研究的课题）进行长期跟踪，一旦发现最新的、有参考价值的资源，就主动推送给科研用户。利用这种积累与服务机制，当在用户提出需求时，虚拟数字图书馆即可作出最为及时的响应。

6. 共享性增强

网格的目的是共享，这种共享已经不再是简单的资源互连和单一使用，而是通过资源互连、资源组合和资源协作来解决用户需要解决的问题，产生具有附加值的新服务、数据、知识等资源，满足用户的新需求。网格把整个因特网整合成为一个巨大的超级计算机，实现网上所有资源的全面连通、全面共享。网格技术的进一步充分应用，能够极大增强数字图书馆的共享性能，从而提高数字图书馆的利用效率。

（1）共享范围——共享主体从内到外、共享对象从单一到综合

首先资源共享主体与客体在范围上进一步扩大。信息资源实现共享的基础条件包括：用户可以通过多种渠道传递信息；在不影响信息再传递与再利用的情况下，多个用户可以同时接收、利用同一内容的信息；用户可以通过多种方式大量复制、加工、整理和利用信息。在 e-Science 出现之前，采用数字化形式来表达信息是保证上述三项基础条件得以实现的根本，对于以数字化资源作为核心资源的数字图书馆而言，信息资源共享是它固有的优点。在 e-Science 的推动下，数字图书馆与科研机构之间以及数字图书馆与数字图书馆之间可以共享的资源将由 e-Science 出现前的信

息资源进一步扩展到设备资源、仪器资源、计算资源、通信资源、专家资源等。无论是共享主体，还是共享客体，其范围都在扩大。

（2）共享质量更高——从信息共享到知识共享

信息共享很大程度上依赖于信息技术和系统来实现，高度发达的技术系统，包括信息的索引、组织控制、查询和传递的技术，可以大大提高信息共享的效率和成果。知识共享可以理解为在知识发展和利用层面上，通过人与人之间建立紧密关系、相互接触和会话，而实现知识资源的复用和知识创新的不断推进。对知识共享来说，单是有技术和系统还远远不够。

信息由信息载体、信息符号及信息内容构成，信息共享建立在信息技术和信息系统之上。得益于各种信息处理技术，例如，信息索引技术、信息组织技术、信息查询与传递技术，信息共享的效率和成果才得以进一步发展。重要的是，信息载体形式的变化、符号系统的转换并不影响信息内容。知识共享是信息共享进一步发展的结果。与信息共享相比，知识共享需要以人与人之间建立紧密关系、相互沟通与交流为基础，依赖于人的思想、价值、文化以及体制、环境，因此，知识共享的实现相对困难。然而，e-Science 环境下，伴随着社会分工的进一步加剧，人们并不希望通过共享原始信息并自行对其进行加工使之上升为知识，而是希望直接获取具有行动指导价值的知识，知识需求的欲望进一步刺激了知识共享的发展。在信息共享上表现出色的虚拟数字图书馆将紧紧抓住 e-Science 带来的发展契机，利用网络可以直接创造知识的特点与优势，全力推进知识共享。借助各种更为实时化、生成图像更为清晰化的工具，虚拟组织内部的成员之间、不同虚拟组织的成员之间的相互信任程度因为有了协议的保证从而使得它比面对面的信任程度更强。这种知识创造的便利性、用户之间信任的充分性都为开展知识共享提供了非常有利的前提条件。

除了启动 DILIGENT 项目，欧洲已有多家组织与研究机构开始致力于网格技术与数字图书馆技术结合的研究。DELOS 网络组织实施了一个联合的计划，其目的在于，关注欧洲地区一些在数字图书馆研究方面卓有成效的团队，并对这些研究团队的研究工作实行整合与协调。目前，部分研究团队已将研究目标定位为发展 e-Science 环境下的数字图书馆技术。这种技术将支撑任何用户在任何地方的任何时刻以一种友好的、多模式可选的、高效的方式去存取人类知识。

欧洲虚拟数字图书馆原型项目 DILIGENT 的启动表明，以网格为基础的 e-Science 将直接影响知识创新与科学研究的模式，并将成为下一代数字图书馆的工作环境，这将是一种必然。一方面，e-Science 不仅给数字图书馆带来更多更为先进的技术，而且进一步扩大数字图书馆可支配资源的范围，提升数字图书馆的服务能力；另一方面，数字图书馆将会以应对 e-Science 带来的各种复杂信息需求、知识需求为发展重点，将自身的各项业务置身于整个 e-Science 环境中加以考虑，并逐步向虚拟数字图书馆发展。

总之，DILIGENT 以 EGEE 为支撑，底层采用与 OGSA 兼容的构件组建，建设

理念与服务机制符合 e-Science 环境下数字图书馆的运作模式。DILIGENT 试验平台建立在网格技术与数字图书馆技术加以集成的基础之上。这种集成将为部署下一代 e-Science 知识基础设施打下坚实的根基。事实上，尽管 e-Science 目前还处于起步阶段，e-Science 环境下数字图书馆建设还不全面与深入，但是，密切关注欧洲数字图书馆的发展趋势，借鉴其虚拟数字图书馆的建设模式与发展策略，将有利于图书情报界更好地在 e-Science 环境下建设和利用数字图书馆。

参 考 文 献

- [1] 徐志伟, 李伟.织女星网络的体系结构研究 [J].计算机研究与发展.2002(8):923-929.
- [2] 刘鹏.我国网格研究现状 [EB/OL].
<http://grid.hust.edu.cn/cgcl/documents/Grid/The%20research%20conditions%20of%20Grid%20in%20China.pdf>. [2007-6-11]
- [3] 杜渐.网格技术等待商业应用突破 [EB/OL].2006-3-10.
<http://www.istis.sh.cn/list/list.asp?id=2657>. [2007-6-11]
- [4] 中国国家网络首页 [EB/OL]. <http://www.cngrid.org/index.htm>. [2007-6-11]
- [5] 陈萍.北京大学网格计算科研与应用 [EB/OL].2005-11.
http://conference.jucc.edu.hk/notes/2.17_Professor_Chen_Ping.pdf. [2007-6-11]
- [6] 中国国家网络(China National Grid)[EB/OL].http://www.cngrid.org/cn_introduce.htm. [2007-6-11]
- [7] 晓齐.网格: IT 技术新热点[J].信息安全.2003(3):59-59
- [8] 刘路沙.中国: 不能错过网格的机遇 [EB/OL].http://www.ncic.ac.cn/news/news_08.htm. [2007-6-11]
- [9] 钱德沛.中国国家网格进展 [EB/OL].2004-4-29.
<http://www.chinagrid.net/grid/paperppt/bigfile/Conference/IDG2004Spring/K1.ppt>. [2007-6-11]
- [10] 中国科技传播网 [EB/OL].<http://book.netpub.com.cn/rssfeed.asp?cid=2>. [2007-6-11]
- [11] 刘斌.中国国家网格正式开通 [EB/OL].2005-12-23.
<http://www.biotech.org.cn/news/news/show.php?id=29484>. [EB/OL]
- [12] 中国国家网格运行管理支持中心. 中国国家网格参加法国 Grid PlugTests 大赛 [EB/OL].2005-10-21.
http://www.cngrid.org/02_news/news/cn_news_news_051021.htm. [2007-6-11]
- [13] xuhaidong.OMII-China delegation visit the University of Southampton, United Kingdom [EB/OL]. 2006-07-14.<http://www.crown.org.cn/en/news/news.jsp?newsID=92>. [2007-6-11]
- [14] 上海交通大学.谭宗仁博士受聘上海网格总体组顾问[EB/OL].2004-6-25.
<http://www.sjtu.edu.cn/newsnet/newsdisplay.php?id=1029>. [2007-6-11]
- [15] Advances in ShanghaiGrid (in Chinese)[EB/OL].
<http://unpan1.un.org/intradoc/groups/public/documents/APCITY/UNPAN016937.pdf>. [2007-6-11]
- [16] 中国教育科研网格.ChinaGrid 简介[EB/OL].
<http://www.chinagrid.edu.cn/chinagrid/index.jsp>. [2007-6-11]
- [17] hp 中国.ChinaGrid 计划全面启动 [EB/OL].
http://www.hp.com.cn/enterprise/success/server/051116_02.asp. [2007-6-11]
- [18] 金海. ChinaGrid 建设目标 [EB/OL].2005-5-30.
<http://www.cutech.edu.cn/jiaoyuxinxihua/000132.asp>. [2007-6-11]

- [19] 金海.ChinaGrid 建设目标: 最大 最先进 最实用 [EB/OL].2007-05-31.
http://www.media.edu.cn/zhuan_jia_yuan_di_5194/20060517/t20060517_179138.shtml. [EB/OL]
- [20] ChinaGrid 公共平台建设 [EB/OL].<http://grid.hust.edu.cn/platform/html/chinagrid.html>. [2007-6-11]
- [21] 基于广域网络的科研环境 [EB/OL].<http://www.crown.org.cn/>. [2007-6-11]
- [22] 关于 CROWN [EB/OL]. http://www.crown.org.cn/aboutus/about_us.jsp. [2007-6-11]
- [23] 怀进鹏. 网络计算环境中件技术研究 2005 年度检查报告-总体进展[EB/OL].2005-12-19.
<http://www.crown.org.cn/download/download.jsp?iDownloadType=2>. [2007-6-11]
- [24] 怀进鹏. 网络计算环境综合试验平台—北航进展汇报 [EB/OL].2005-12-19.
<http://www.crown.org.cn/download/download.jsp?iDownloadType=2>. [2007-6-11]
- [25] 刘恕.让地质空间信息自由流动 [EB/OL].
http://www.863.org.cn/863_105/news/news_infotech/200507150001.html. [2007-6-12]
- [26] 空间信息网格概述[EB/OL].
<http://www.nudt.edu.cn/newweb/intercommunion/whatissig.htm>. [2007-6-12]
- [27] 863 计划信息获取与处理技术主题组织部分优秀成果参加第 18 届国际卫星对地观测委员会 (CEOS) 全会展览会 [EB/OL].2004-11.http://www.most.gov.cn/kjbgz/200411/t20041126_17530.htm. [2007-6-12]
- [28] 金江军.网格技术在地球信息科学中的应用 [EB/OL].2004-2-25.
http://industry.ccidnet.com/art/322/20040225/92292_1.html. [2007-6-12]
- [29] SIG 空间信息网格 [EB/OL].http://grid.hust.edu.cn/article_63.html. [2007-6-12]
- [30] <http://www.lehigh.edu/~bad0/fafner.html> [2005-11-29]
- [31] <http://www.nitrd.gov/pubs/bluebooks/1997/i-way.html> (2005-11-29)
- [32] <http://www.accessgrid.org/> (2005-11-29)
- [33] <http://www.teragrid.org/> (2005-11-29)
- [34] <http://www.griphyn.org/index.php> (2005-11-30)
- [35] <http://www.ppdg.net/> (2005-11-30)
- [36] <http://www.usatlas.bnl.gov/> (2005-11-30)
- [37] <http://cmcs.ca.sandia.gov/> (2005-11-30)
- [38] <http://www.baldridge.unizh.ch/resurgence/index.html> (2005-11-30)
- [39] <http://www.us-vo.org/> (2005-11-30)
- [40] <http://www.phy.ornl.gov/tsi/> (2005-11-30)
- [41] <http://it.nees.org/> (2005-11-30)
- [42] <http://www.ccsn.ucar.edu/index.html> (2005-11-30)
- [43] <http://www.nbirn.net/index.htm> (2005-11-30)
- [44] <http://seek.ecoinformatics.org/> (2005-11-30)
- [45] <http://www.ipg.nasa.gov/> (2005-11-30)
- [46] <http://www-unix.mcs.anl.gov/scidac-tops/> (2005-11-30)
- [47] <http://www.nsf-middleware.org/default.aspx> (2005-11-30)

- [48] <http://www.nsf.gov/> (2005-11-30)
- [49] <http://www.scidac.org/index.html> (2005-12-1)
- [50] <http://www.doecollaboratory.org/homepage.html> (2005-12-1)
- [51] <https://www.earthsystemgrid.org/> (2005-12-1)
- [52] <http://www.fusiongrid.org/> (2005-12-1)
- [53] <http://vdt.cs.wisc.edu/index.html> (2005-12-2)
- [54] <http://www.star.bnl.gov/> (2005-12-3)
- [55] <http://www.pragma-grid.net/> (2005-12-3)
- [56] <http://www.slac.stanford.edu/BFROOT/> (2005-12-3)
- [57] e-Science Project Mapping. http://www.nesc.ac.uk/projects/project_mapping.html (2005-12-3)
- [58] <http://www.ultralighthomepage.com/> (2005-12-4)
- [59] DOE Science Grid: Enabling and Deploying the SciDAC Collaboratory Software Environment. http://www.doesciencegrid.org/Grid/Grid/papers/DOE_Science_Grid_Collaboratory_Pilot_Proposal_03_14.nobudget.pdf (2005-12-4)
- [60] Software Components for Grid Systems And Applications. <http://www-unix.grids-center.org/r6/ecosystem/> (2005-12-4)
- [61] <http://www.lsc-group.phys.uwm.edu/LDR/> (2005-12-4)
- [62] <http://www.nsf.gov/dir/index.jsp?org=OCI> (2005-12-5)
- [63] GRIDSTART Project: An Executive Summary of the Inventory and Roadmap of the FP5 IST Grid Projects
- [64] EU Grid Projects Achievements: As funded by the EU' s Fifth Framework Programme (FP5). Compiled and edited by the GRIDSTART Consortium
- [65] GridCoord: ERA Pilot on a co-ordinated Europe-wide Initiative in Grid Research. Survey of Funding Bodies and Industry
- [66] GRIDSTART Project: IST Grid Projects' Inventory and Roadmap
- [67] <http://www.escidoc-project.de/en/index.html>
- [68] http://europa.eu.int/comm/research/headlines/news/article_04_10_21_en.html
- [69] http://inside.hlr.de/htm/Edition_01_04/article_04.htm
- [70] http://www.fiz-karlsruhe.de/pressroom/pilot-mpg_en.html
- [71] <http://www.nesc.ac.uk/nesc/define.html>
- [72] <http://www.crossgrid.org/>
- [73] <http://www.hlr.de/organization/pds/projects/damien/>
- [74] <http://www.cogproject.org/>
- [75] <http://www.asp-bp.org>
- [76] <http://www.mssl.ucl.ac.uk/grid/egso/>
- [77] <http://datatag.web.cern.ch/datatag/>
- [78] <http://www.bio-grid.net/index.jsp>

- [79] <http://www.gemss.de>
- [80] <http://www.openmolgrid.org/>
- [81] <http://www.datamininggrid.org/>
- [82] <http://eu-datagrid.web.cern.ch/eu-datagrid/>
- [83] <http://www.unigrids.org/>
- [84] <http://www.lege-wg.org/>
- [85] <http://www.ib-ia.com/websi/>
- [86] <http://www.cern.ch/ttdb/Projects/mammogrid>
- [87] <http://www.coregrid.net/>
- [88] <http://www.kwfgrid.net/main.asp>
- [89] <http://www.hpc4u.org/public/>
- [90] <http://www.ontogrid.net/>
- [91] <http://www.nextgrid.org/>
- [92] <http://www.inteligrid.com/>
- [93] <http://twiki.gridprovenance.org/bin/view/Provenance/>
- [94] <http://www.simdat.org>
- [95] <http://www.deisa.org/>
- [96] <http://www.dcs.bbk.ac.uk/selene/>
- [97] <http://www.grace-ist.org/>
- [98] <http://www.gria.org/>
- [99] <http://www.gridlab.org/>
- [100] <http://www.euro-vo.org/>
- [101] <http://www.eurogrid.org/>
- [102] <http://egee-intranet.web.cern.ch/egee-intranet/gateway.html>
- [103] <http://www.gridstart.org/>
- [104] <http://www.unizar.es/flowgrid/>
- [105] <http://asg-platform.org/cgi-bin/twiki/view/Public>
- [106] <http://www.gridatasia.net/>
- [107] <http://www.akogrimo.org/>
- [108] <http://www.grid-interoperability.org/>
- [109] <http://www.hum.ku.dk/moses/index2.htm>
- [110] <http://www.gridcoord.org>
- [111] <http://www.eu-grasp.net/>
- [112] <http://www.bast.net.cn/bjkipzc/kjqy/xxkx/35783.shtml>
- [113] http://grid.hust.edu.cn/article_63.html
- [114] <http://www.chinagrid.net/dvnews/show.aspx?id=355&cid=14>
- [115] http://www.zjgg.com.cn/public/news_1.jsp?id=753

- [116] <http://e.chinabyte.com/473/2038973.shtml>
- [117] http://www.cngrid.org/08_maps/jiedian_zhongkeyuan.htm
- [118] http://www.cngrid.org/08_maps/jiedian_shanghai.htm
- [119] <http://www.gridcenter.or.kr>
- [120] http://www.collab-ogce.org/nmi/portal/user/anon/js_pane/P-fc75d05650-10001
- [121] <http://www.twaren.net/chinese/about/>
- [122] <http://webgis.nchc.org.tw/kingnew/ans.html>
- [123] http://www.intellectcube.jp/~kenji.nihei/NAREGI_ENG//papers/index_e.html#top
- [124] <http://www.ksc.re.kr/eng/user/user1.htm>
- [125] <http://www.crown.org.cn/>
- [126] 钱德沛. 发展 e-Science 促进科研信息化[EB/OL].2005-10-24
http://www.stdaily.com/gb/stdaily/2005-10/24/content_446421.htm. [2005-11-28].
- [127] Liz Lyon. e-Research: trends, requirements and challenges[EB/OL].
<http://www.ukoln.ac.uk/ukoln/staff/e.j.lyon/crossrc-ict-may04.ppt>. [2005-11-2]
- [128] E-Science Evaluation Report – 2004. <http://www.rcuk.ac.uk/escience/documents/evaluationreport2004.pdf>
[2005-11-12]. e-Science Project Mapping [EB/OL]. 2005-10-7
http://www.nesc.ac.uk/projects/project_mapping.html. [2005-10-08]
- [129] Rob Allan. Service Classes for the e-FER [EB/OL]. <http://www.grids.ac.uk/Papers/Classes/classes.html> [2005-11-22].
- [130] DAME [EB/OL]. <http://www.cs.york.ac.uk/dame/>. [2005-10-8].
- [131] Sarfraz A Nadeem, Towards the Design and Implementation of the DAME prototype: OGSA Compliant Grid Services on the White Rose Grid [EB/OL]. <http://www.cs.york.ac.uk/dame/DAMEopendayXTO.ppt>. [2005-9-17].
- [132] 黄理灿, 刘元安. e-Science、网格及其可扩展性体系结构 [M]. 北京: 北京邮电大学出版社, 2005: 6.
- [133] ATF 2001 - August 2003 [EB/OL]. <http://www.nesc.ac.uk/teams/atf/01-03/>. [2005-11-22]
- [134] Grid Network Team [EB/OL]. <http://www.nesc.ac.uk/teams/gnt.html>. [2005-11-25]
- [135] Database Access and Integration Services Working Group [EB/OL]. <http://www.cs.man.ac.uk/grid-db/>. [2005-11-30]
- [136] UK e-Science Architecture Task Force [EB/OL]. <http://www.nesc.ac.uk/teams/atf.html>. [2005-11-30]
- [137] The UK Grid Engineering Task Force [EB/OL]. <http://tyne.dl.ac.uk/ETF/index.shtml> [2005-11-30].
- [138] Security Task Force [EB/OL]. <http://www.nesc.ac.uk/teams/stf/> [2005-11-30].
- [139] Usability Task Force (UTF) [EB/OL]. <http://www.nesc.ac.uk/teams/> [2005-11-30].
- [146] UK e-Science User Group [EB/OL]. <http://www.nesc.ac.uk/teams/usergroup/> [2005-11-30].

附录A 美国e-Science相关项目列表

ID	项 目 名 称	URL
1	A High-Performance Data Grid Toolkit: Enabling Technology for Wide Area Data-Intensive Applications	http://www-fp.mcs.anl.gov/dsl/scidac/datagrid/
2	A Powerful Distributed Cyberinfrastructure to Support	http://www.optiputer.net/
3	A Scalable and Secure Peer-to-Peer Information Sharing Tool	http://dsd.lbl.gov/P2P/file-share/
4	Access Grid	http://www.accessgrid.org/
5	Advanced Computing for 21st Century Accelerator Science and Technology	http://www.scidac.org/HENP/HENP_accelerator.html
6	Bandwidth Estimation: Measurement Methodologies and Applications	http://www.caida.org/projects/bwest/index.xml
7	Biomedical Informatics Research Network	http://www.nbirn.net/index.htm
8	BNL grid site Authentication Authorization and Accounting page	http://grid.racf.bnl.gov/siteAAA/index.html
9	Center for Component Technology for Terascale Simulation Software	http://www.scidac.org/ASCR/ASCR_CCTSS.html
10	Collaboratory for Multi-Scale Chemical Science	http://cmcs.ca.sandia.gov/
11	Commodity Grid Kits	http://www.cogkit.org/
12	Common Instrument Middleware Architecture	http://www.instrumentmiddleware.org/metadot/index.pl?iid=2119&isa=Category
13	Community Climate System Model	http://www.ccsm.ucar.edu/index.html
14	Community Services	http://community-services.cs.umn.edu/
15	Condor	http://www.cs.wisc.edu/condor/
16	Continuous Dynamic Grid Adaptation	http://www.pircs.iastate.edu/people/Gutowski/DOE_CCPP.html
17	Coupled Colorado State Model	http://kiwi.atmos.colostate.edu/DOE_Cooperative_Agreement/
18	Cryptographic Foundations for New Generation Distributed Systems	http://dsd.lbl.gov/Projects/OPKeyX/
19	DataCutter	http://www.datacutter.org/
20	Decadal Regional Climate Studies and Applications with Variable-Resolution GCMs Using Advanced Numerical Techniques	http://www.scidac.org/BER/BER_VR_GCM.html

ID	项 目 名 称	URL
21	Development of an Atmospheric Climate Model with Self-Adapting Grid and Physics	http://www.scidac.org/BER/BER_SAGrid.html
22	Distributed Monitoring Framework	http://dsd.lbl.gov/DMF/
23	Distributed Security Architectures	http://dsd.lbl.gov/security/DistSecArch.html
24	Distributed Systems Department	http://dsd.lbl.gov/
25	DOE Science Grid	http://www.doeSciencegrid.org/
26	Earth Observing System	http://eosps.gsfc.nasa.gov/index.php
27	Earth System Grid	https://www.earthsystemgrid.org/
28	Earth System Modeling Framework	http://www.esmf.ucar.edu/
29	eServices Infrastructure for Collaborative Science	http://www-unix.mcs.anl.gov/eservices/
30	GeoSciences Network	http://www.geogrid.org/
31	Globus	http://www.globus.org/
32	Grid Application Development Software Project	http://www.hipersoft.rice.edu/grads/
33	Grid Physics Network	http://www.griphyn.org/index.php
34	Grid Physics Network in ATLAS	http://www.usatlas.bnl.gov/computing/grid/griphyn/
35	Grid Research Integration Deployment and Support Center	http://grids-center.org/
36	Grid User Management System	http://grid.racf.bnl.gov/GUMS/
37	GridShib	http://gridshib.globus.org/
38	INCITE – Edge-based Traffic Processing and Inference for High-Performance Networks	http://www-ece.rice.edu/INCITE/
39	Information Power Grid	http://www.ipg.nasa.gov/
40	International Virtual Data Grid Labortory	http://www.ivdgl.org/
41	IQ-ECho : Interoperability and Quality of Service Across Heterogeneous Hardware/Software Platforms	http://www.cc.gatech.edu/systems/projects/IQECho/
42	iVDGL Tier2 Operations for ATLAS	http://tier2.iu.edu/index.shtml
43	Kepler Project	http://kepler-project.org/
44	Linked Environments for Atomspheric Discovery	http://lead.ou.edu/
45	Long Term Ecological Research Network	http://www.lternet.edu/
46	Manager for Grid-based Data	http://www.atlasgrid.bnl.gov/magda/info/
47	Middleware Technology to Support Science Portals: a Gateway to the Grid	http://www.scidac.org/ASCR/ASCR_MTSSports.html
48	Middleware to Support Group to Group Collaboration	http://www-unix.mcs.anl.gov/fl/research/G2G/
49	National Fusion Grid	http://www.fusiongrid.org/
50	National Laboratory for Advanced Data Research	http://www.nladr.net/
51	National Science Foundation Middleware Initiative	http://www.nsf-middleware.org/default.aspx

ID	项 目 名 称	URL
52	Network for Earthquake Engineering Simulation	http://it.nees.org/
53	Networked Object-Based EnVironment for Analysis	http://www.usatlas.bnl.gov/computing/nova/
54	NMI-EDIT	http://www.nmi-edit.org/index.cfm
55	Notebook Project	http://www.notebookproject.org/
56	Open Grid Computing Environments Collaboratory	http://www.collab-ogce.org/nmi/index.jsp
57	Open Science Grid	http://www.openSciencegrid.org/index.php?option=com_frontpage&elMenu=Home
58	Open-Source for a Network Data Access Protocol	http://www.opendap.org/index.html
59	Optimizing Performance and Enhancing Functionality of Distributed Applications using Logistical Networking	http://loci.cs.utk.edu/scidac/
60	Particle Physics Data Grid	http://www.ppdg.net/
61	Particle Physics Data Grid (PPDG) in ATLAS	http://www.usatlas.bnl.gov/computing/grid/ppdg/
62	Performance Evaluation Research Center	http://perc.nersc.gov/main.htm
63	Performance Inside	http://www-unix.mcs.anl.gov/~schopf/PerfInside/
64	Pervasive Collaborative Computing Environment	http://dsd.lbl.gov/Collaboratories/pcce.html
65	Planning for Execution in Grids	http://pegasus.isi.edu/
66	Plasma Microturbulence Project	http://www.scidac.org/FES/FES_PMP.html
67	Portal Web Services: Support of DOE SciDAC Collaboratories	http://www.doeportals.org/
68	pyGridWare: Python Web Services Resource Framework	http://dsd.lbl.gov/gtg/projects/pyGridWare/
69	Python Globus	http://dsd.lbl.gov/gtg/projects/pyGlobus/
70	Real-time Observatories Applications and Data management Network	http://roadnet.ucsd.edu/
71	Reliable and Secure Group Communication	http://dsd.lbl.gov/GroupComm/
72	Research Surge Enabled by Cyberinfrastructure	http://www.baldrige.unizh.ch/resurgence/index.html
73	Scalable Systems Software for Terascale Computer Centers	http://www.scidac.org/ScalableSystems/
74	Science Environment for Ecological Knowledge	http://seek.ecoinformatics.org/
75	Scientific Annotation Middleware	http://collaboratory.emsl.pnl.gov/docs/collab/sam/index.html
76	Scientific Data Management Center	http://sdm.lbl.gov/sdmcenter/
77	Scientific Discovery through Advanced Computing	http://www.scidac.org/index.html
78	SDSC Storage Resource Broker	http://www.sdsc.edu/srb/
79	Security and Policy for Group Collaboration	http://www-fp.mcs.anl.gov/dsl/scidac/security/
80	Sequential data Access via Meta-data	http://d0db-prd.fnal.gov/sam/
81	Storage Resource Broker	http://www.sdsc.edu/srb/

ID	项 目 名 称	URL
82	Storage Resource Management Middleware Project	http://sdm.lbl.gov/indexproj.php?ProjectID=SRM
83	Supernova Science Center	http://www.supersci.org/
84	TeraGrid	http://www.teragrid.org/
85	Terascale Optimal PDE Simulations	http://www-unix.mcs.anl.gov/scidac-tops/
86	Terascale Simulation Tools and Technology Center	http://www.tstt-scidac.org/
87	TeraScale Supernova Initiative	http://www.phy.ornl.gov/tsi/
88	U.S. ATLAS	http://www.usatlas.bnl.gov/
89	U.S. Compact Muon Solenoid Software and Computing	http://www.uscms.org/SoftwareComputing/index.html
90	U.S. Department of Energy's National Collaboratories program	http://www.doeccollaboratory.org/
91	U.S. National Virtual Observatory	http://www.us-vo.org/
92	Variable-Resolution Models for Regional Climate Studies	http://www.osti.gov/scidac/updates2004/ber_14.html
93	Virtual Data Toolkit	http://vdt.cs.wisc.edu/index.html
94	Virtual Grid Application Development Software	http://vgrads.rice.edu/
95	VO Privilege Project	http://computing.fnal.gov/docs/products/voprivilege/index.html
96	A Market for Computational Services	http://www.lesc.ic.ac.uk/markets/
97	BaBar	http://www.slac.stanford.edu/BFROOT/
98	Computing Systems of Microbial Interactions and Communications on a GRID	http://www.nesc.ac.uk/action/projects/project_action.cfm?Title=130
99	Integrative Biology	http://www.integrativebiology.ox.ac.uk/
100	IPv4 + IPv6 Performance and QoS	http://grid.ucl.ac.uk/46PaQ.html
101	Managed Bandwidth - Next Generation	http://www.grid.ucl.ac.uk/MB-NG
102	Messaging and Notifications for Grid Organisations	http://www.wesc.ac.uk/projects/mango/
103	MyGrid	http://www.mygrid.org.uk/
104	Open Grid Services Architecture - Data Access and Integration Services	http://www.ogsadai.org.uk/
105	Pacific Rim Applications and Grid Middleware Assembly	http://www.pragma-grid.net/
106	RAVE: A Resource-Aware Visualisation Environment	http://www.nesc.ac.uk/action/projects/project_action.cfm?Title=101
107	Service Composition for Data Exploration in the Virtual Observatory (SC4DEVO)	http://gow.epsrc.ac.uk/ViewGrant.ASPx?Grant=GR/S85023/01&bannerlink=Programme%20support
108	SimGrid	http://simgrid.gforge.inria.fr/
109	Solenoidal Tracker At RHIC experiment	http://www.star.bnl.gov/
110	The Screensaver Lifesaver Project	http://e-Science.ox.ac.uk/public/eprojects/cancer/index.xml?style=printable
111	VideoWorks for the Grid	http://www.videoworks.ac.uk

附录B 欧盟第五框架计划和第六 框架计划资助的项目

项目名称	参与机构 数目	参与国 家数目	项目负责人 所在国家	资助额度 (单位: 欧元)	开始年份	项目期限 (单位: 月)	欧洲以 外参与国家
eurogrid	10	6	德国	2070000	2000.11	39	
damien	5	3	德国	1230000	2001.1	34	
datagrid	15	10	CERN	9800000	2001.1	36	
gria	6	4	英国	2020000	2001.12	36	
Gridlab	14	10	波兰	5085998	2002.1	40	美国
Grip	6	3	德国	1340000	2002.1	26	
Datetag	8	4	CERN	3980000	2002.1	24	
Selene	5	4	英国	N/A	2002.11	15	
crossgrid	21	11	波兰	4841582	2002.3	38	
Egso	11	5	英国	2400000	2002.3	36	美国
Grasp	8	6	西班牙	1960000	2002.4	34	
gridstart	18	4	英国	1450000	2002.4	36	
Websi	6	4	西班牙	1799987	2002.5	29	
asp-bp	36	8	意大利	3490000	2002.5	24	
LeGE-WG	23	9	英国	715580	2002.8	24	
Biogrid	6	4	德国	834000	2002.9	27	
flowgrid	6	5	西班牙	1100000	2002.9	28	
Gems	10	5	英国	2630000	2002.9	30	
Grace	6	5	意大利	1890000	2002.9	30	
mammogrid	6	2	CERN	1900000	2002.9	36	
openmoigrid	5	5	德国	1990000	2002.9	30	
cog	8	5	希腊	1060000	2002.9	18	
Avo	6	3	德国	5000000	2001.11	36	
Moses	7	4	意大利	N/A	2002.2	30	
egee	71	27	N/A	32000000	2004.4	24	美国、 俄罗斯
deisa	11	7	法国	N/A	2004.5	N/A	
hpc4u	8	6	法国	1700246	2004.6	36	
Akogrimo	13	9	西班牙	7000000	2004.9	36	
coregrid	42	18	法国	8200000	2004.9	48	智利

续表

项目名称	参与机构 数目	参与国 家数目	项目负责人 所在国家	资助额度 (单位: 欧元)	开始年份	项目期限 (单位: 月)	欧洲以 外参与国家
datamininggrid	5	4	英国	883000	2004.9	36	
gridcoord	13	10	意大利	960000	2004.9	18	
inteligrid	9	7	斯洛文尼亚	2122000	2004.9	30	
kwfgrid	6	6	德国	1746822	2004.9	30	
nextgrid	22	7	英国	11000000	2004.9	48	
ontogrid	8	4	西班牙	2638940	2004.9	36	
provenance	6	4	英国	1981996	2004.9	24	
simdat	25	6	德国	11000000	2004.9	48	国际组织
unigrids	7	4	德国	1916162	2004.9	24	
asg	22	7	N/A	N/A	2004.9	24	澳大利亚
gridasia	7	5	法国	300000	2005.4	18	中国、韩国

注: 白色底为欧盟第五框架资助项目; 将表中深色部分底为欧盟第六框架资助项目; N/A 表示该项数据未知。

附录C 欧盟第五框架计划下网格 项目研究网格技术层次

网格层次/ 项目名称	基础设施	中间件	开发环境和工具	应用和入口
AVO	●	●		●
CrossGrid	●	●	●	●
DAMIEN		●	●	●
DataGrid	●	●	●	●
DataTAG	●	●		●
EGSO		●	●	●
EUROGRID		●	●	●
GRIA		●		●
GridLab	●	●	●	●
GRIP		●	●	●
BioGrid				●
COG	●	●	●	●
FlowGrid		●	●	●
GEMSS	●	●		●
GRACE		●		●
GRASP	●	●		●
MammoGrid				●
MOSES	●	●		●
OpenMolGRID		●	●	●
SeLeNe		●	●	●

注：浅色底为 FP5 第一阶段项目；深色底为 FP5 第二阶段项目

附录D 欧盟第五框架下网格项目开发的组件

开 发 领 域	开发的组件	涉及的项目
网格入口	<ul style="list-style-type: none">• Genius portal• Earth Observation Portal• AliEn web portal	<ul style="list-style-type: none">• DataGrid• MammoGrid• AliEn
	<ul style="list-style-type: none">• Migrating desktop	<ul style="list-style-type: none">• CrossGrid
	<ul style="list-style-type: none">• GridSphere framework, portlets and portlet development environment	<ul style="list-style-type: none">• GridLab
	<ul style="list-style-type: none">• UNICORE client along with plug-ins	<ul style="list-style-type: none">• EuroGrid• GRIP• OpenMolGRID
网格应用	<ul style="list-style-type: none">• Biotechnology applications	<ul style="list-style-type: none">• EuroGrid• GRIP• OpenMolGrid• BioGrid
	<ul style="list-style-type: none">• Cactus-based applications (general purpose framework, large user community in astrophysics)	<ul style="list-style-type: none">• GridLab
	<ul style="list-style-type: none">• Particle physics applications	<ul style="list-style-type: none">• DataGrid• CrossGrid
	<ul style="list-style-type: none">• Engineering applications	<ul style="list-style-type: none">• DAMIEN• EuroGrid• GRIA• FlowGrid
	<ul style="list-style-type: none">• Medical applications	<ul style="list-style-type: none">• CrossGrid• GEMSS• MammoGrid
应用环境 和工具	<ul style="list-style-type: none">• Grid Application Toolkit (GAT)	<ul style="list-style-type: none">• GridLab
	<ul style="list-style-type: none">• PACX-MPI	<ul style="list-style-type: none">• DAMIEN – used by CrossGrid
	<ul style="list-style-type: none">• Extended Code Coupling Interface – MpCCI, VAMPIR, and DIMEMAS	<ul style="list-style-type: none">• DAMIEN
	<ul style="list-style-type: none">• UNICORE plug-ins	<ul style="list-style-type: none">• EuroGrid• OpenMolGRID
	<ul style="list-style-type: none">• MARMOT MPI verification tool	<ul style="list-style-type: none">• CrossGrid
资源管理和日程 安排	<ul style="list-style-type: none">• Job submission service (JSS)	<ul style="list-style-type: none">• DataGrid
	<ul style="list-style-type: none">• QoS Manager for MPI applications	<ul style="list-style-type: none">• DAMIEN

续表

开 发 领 域	开发的组件	涉及的项目
	• EuroGrid Resource Broker （based on UNICORE）	• EuroGrid
	• Use of Globus resources through UNICORE	• GRIP
	• Grid （Lab） Resource Management System （GRMS）	• GridLab
信息服务	• GLUE schema	• DataTAG
	• MDS extension	• GridLab
	• iGrid	
监测和绩效	• R-GMA • Fabric Monitoring and Fault Tolerance • Network Cost Estimation Service • Network Monitor Architecture （NMA） • GRM/PROVE	• DataGrid
	• Mercury • Delphi	• GridLab
	• MPI application monitoring, VAMPIR • Performance prediction , DIMEMAS	• DAMIEN
	• Online interactive performance analysis with G-PM • Online monitoring of Grid applications with OCM-G	• CrossGrid
登录	• Logging and Bookkeeping service	• DataGrid
数据管理	• Hierarchical Replica Management	• DataGrid
	• Spitfire	• DataGrid
	• Data Access Prediction	• CrossGrid • DataGrid
安全	• VOMS	• DataGrid
	• Proxy-CA for GSI-proxy generation	• GRIP
	• GridLab Authorisation Service （GAS）	• GridLab
测试平台	• Automatic and periodic tests, and status updates	• GridLab

附录E 中国e-Science相关项目列表

ID	项 目 名 称	项 目 类 型
1	中国教育科研网格	应用研究型
2	图像处理网格	工具研究型
3	生物信息学网格	应用研究型
4	大学课程在线网格	应用研究型
5	计算流体力学网格	应用研究型
6	海量数据处理网格	工具研究型
7	AMS02 数据处理和分析系统	应用研究型
8	大学数字博物馆网格	应用研究型
9	中国气象应用网格	应用研究型
10	资源环境网格	应用研究型
11	航空制造网格	应用研究型
12	科学数据网格	工具研究型
13	新药研发网格	应用研究型
14	生物信息应用网格	应用研究型
15	森林资源与林业生态工程信息应用网格	应用研究型
16	织女星网格	产品研发型
17	中科院计算所知识网格	工具研究型
18	上海高校网格（上海市网格技术 E-Institute）	应用研究型
19	上海高校 e-Institut 计划	应用研究型
20	中欧网格	应用研究型
21	先进计算基础设施北京上海试点工程	试验推广型
22	上海网格	应用研究型
23	国家自然科学基金委网格建设项目（CROWN）	应用研究型
24	863 空间信息网格	应用研究型
25	中国国家网格	应用研究型
26	国家计算网格	应用研究型
27	计算科学 E-研究院	应用研究型
28	免疫学 E-研究院	应用研究型
29	模式生物 E-研究院	应用研究型
30	社会学 E-研究院	应用研究型

附录F 调研e-Science项目技术研究内容

Level 1-Auditing	
安全	
信任	
Level 1-Authentication & Authorisation	
Akenti	Site access management
Attribute Authority service	user account assignment and management
Authorization software for NIMI	User-to-account mapping
Client authentication	Virtual Organization Membership Service
Credential storage	VO Management
Grid Identity Mapping Service	VOMS
Grid Proxy Credentials Manager	X.509 identity certificate
Grid User Management System	航空制造网格
Password authentication	系统安全与认证
Registration process	信息资源安全
Level 1-Brokering	
Resource Broker	
Level 1-Collaboration and Remote Instrument Services	
accessibility of instruments	interactive access
Bidirectional IIOP in MICO	interconnectivity
broadband video-conferencing technology	Interoperability
collaborative applications	LTER Network
collaborative authoring environment	Military network infrastructures
Collaboratory Infrastructure	National and international collaborations
Collaboratory Middleware	NPSS Integration with CORBA-IPG
Collaboratory Pilots	Pervasive Collaborative Computing Environment
Control and Observation in Distributed Environment	platform independent standards
CORBA-IPG Infrastructure	remote visualisation
Delivery of services to remote regions	Satellite Caching
Group to Group Middleware	Virtual peer-learning community
integration grid	合作
Level 1-Co-Scheduling	
Co-ordination	NCSA AGScheduler
Job Description Languages	Scheduler
Job Management and Scheduling	Scheduling Workflow Applications

Level 1-Data Access Services	
Event Data and Distributed Data Access	Storage Resource Manager
OpenDAPg	安全访问
Level 1-Data Cataloguing and Provenance	
Provenance service	
virtual data tracking and catalog	
Level 1-Fault Management	
Bug system	
Fault Tolerance and Resilience	
Level 1-Global Event Services	
CORBA-IPG Event Logger	netarchd
Event and logging Systems	pyGMA
Event Filter	
Level 1-Grid Information Service	
Access Grid Documentation Project	科学数据网格的系统平台
Digital Sky Survey 图像检索网格服务	科学数据网格所必需的中间件软件
GSC (Guide Star Catalog) 星表译码和查询网格服务	生物信息学网格
Information Service Interfaces	示范应用系统??虚拟天文台
Messaging Systems	数据网格技术
MyGrid Context service	网格分类引擎
USNO-A2 星表查询网格服务	网格搜索引擎
基本天文计算服务	信息检索系统
基于网格技术的数据库检索	银河系铁元素丰度梯度统计分析
科学数据库中大量分布式异构数据资源的有效共享	知识网格服务
科学数据网格	中国科学院科学数据库
Level 1-Monitoring	
ACDC Grid Dashboard	NMI Sensor Toolkit
Analysis monitoring module	PRAGMA Bandwith
ApGrid	Rocks Network Grid Report
Athena Auditor services	sensor management tools
Cflowd	service catalogs
Distributed Monitoring Framework	status reporting
Grid Monitoring Architecture	Test and Monitoring Service
GridCat	ThaiGrid
GridView	visualizing monitoring data
Job Status Tools	互联互通和协同工作
MAGNET User-Space Environment	林业信息资源共享
MonALISA	森林资源监测与分析

Monitoring Apparatus for General kerNel-Event Tracing	森林资源与林业生态工程网络
MRTG	退耕还林工程管理
NetLogger	远程设备监测
NetLogger Toolkit	支持我国森林资源和林业生态工程的管理与决策
Network Characterization Service	
Traffic Information-Collecting Kernel with Exact Timing	
Level 1-Resource Access Services	
adaptive resource providers	MusicGrid
ARP selection	Network Data Access Protocol
Collaborative Content Creation Lab	peer-to-peer learning object sharing
data transport service	policy-based caching
Digital Rights Management	Portal for Online Objects in Learning
Dynamic caching	pyGridWare
educational resources	Reliable replica management services
EduSource	Repository Content Development
File transfer services	Resource Browsers
file-sharing system	resource management
Firefish	SciShare
identity and access management	SportWeb Online Education
Language Learning Environment and Resource Network	Storage Efficient Access
LOGIC Case-Study Generator	resource management system
MuseEdit	
Level 1-Security and Privacy	
Bro: Network Intrusion Detection	secure attribute sharing
cryptographic primitives	SGL
IETF' s Proxy	Totem reliable multicast protocol
InterGroup Protocol	安全性
Network security	计算资源资源共享
OPKeyX	可靠性
PKI standards	可扩展性
Reliable and Secure Group Communication	数据资源共享
Level 2-Collaboration Toolkits	
Access Grid Toolkit	XMPP text-based messaging component
Collaborative Computing	互操作工具包
Collaborative infrastructure	FLOWGRID Toolkit
condor middleware	Grid portals
data sharing	Grid Remote Procedure Call (GridRPC)
Distance Coaching System	intelligent middleware system

distributed collaborations	LearnCanada
distributed computing resources	Pan-Canadian Health Informatics Collaboratory
easy-to-use interface	Partnerships for Learning Innovation and Technology
ELLnet	personal version of the Access
eminerals virtual	programming middleware
Virtual Data Toolkit	SMS
The Development of an Integrated e-Learning Tool for Case-Based Learning	
University Collaborative Communities for e-Learning Adoption	
Level 2-Data Management Services	
Automatic Resource Characterization	Nux
Catalog interface	OGSA-WebDB
Catalogs	Ontology Based Information Service (Ont-IS)
Collective Data Management	real-time data management
data accessibility	Reliable File Transfer
Data and Storage management	Replica Location Service
Data Assimilation Applications	Robust File Replication
Data Catalog	Scientific Data Management
data grid	Storage Management
Data Grid Management System	Storage resource management
Data Movement	uniform view
Data Repository	Virtual data
Data Transfer Management	多数据的集成与分析
DataGrids	分布式计算
Digital Library	分布式数据存储与共享服务
Distributed File System	航空制造网格
eI-SSB	集成计算
Grand Challenge Architecture	解决单一数据源问题
Grid filesystem	企业相互协作
Grid Information services	数据版本不一致
Grid Resource Configuration Generation	数据存储
GridFTP	数据访问
GridHSM	数据服务
heterogeneous database resources	数据管理
high performance data movement	数据集成
integrate various scientific databases	数据库
Management of data	特征抽取
Management of jobs	异地分布式数据管理和共享模式

metadata-based data management technologies	属性数据
MIX: XQuery	
Level 2-Data Mining and Analysis Services	
Analysis catalog	集成应用
Analysis daemon	决策树
CVS code repository	模型挖掘
Datacutter	数据仓库
Distributed Analysis	数据分析
Dynamically loaded applications	数据界面和服务
filter-stream programing	数据库技术
Kepler	数据挖掘方法
Offline control framework	算法数据源
ROOT analysis	图像特征抽取
web 数据库	文本挖掘
workflow管理工具	文本挖掘界面和组件
集成系统	系统设计
Level 2-Grid Administration Services	
Condor	PIG
Dynamic Accounting System	SRB
Grid computing	XML/CML
MINIGRID	
Level 2-Instrument Management Services	
航空制造网格	硬件资源共享服务
降低企业成本	硬件资源进行合理的调度与管理
缩短研制周期	远程设备监测
虚拟计算中心	远程设备控制
遗传算法计算模型体系	整合集团计算机资源
Level 2-Market Economy Services	
合作	
会计	
支付	
Level 2-Patterns and Program Models	
Accelerator Modeling	pyGridWare
Automatic Construction of Performance Models	航空制造网格
Common Component Architecture	企业内部资源的有效管理和利用
Parallel Software Design Model	提高协同工作效率
performance-enabled NMI software	提供建模和分析工具
PyGlobus	优化工作设计流程

Level 2-Virtual or Immersive Environment Services	
E-LIVE	Simu Status
Grid Modeling, Simulation, and Benchmarking	Virtual Grid
GSTAR	Virtual Veterinary Medicine Learning Community
Running Pythia	虚拟组织
Level 2-Visualisation Services	
4-D evolution of the North American lithosphere	Visualization Analytics
GeeViz	参考模型
Growl	数据可视化
Oilfield Simulation and Siesmic Visualization	
Level 2-Workflow Services	
BPEL workflow tools	workflow management
Collect and document experiment practices	workflow management system
dynamic workflow orchestration	workload management system
Karajan Workflow Engine	存储
Kepler alpha	合作
mapping of complex scientific workflows onto the grid	基于知识的工作流系统
Portlets to Access Application Factory Services	设计优化
Ptolemy II system	制定
scientific workflow system	最优化配置
SDSC Matrix	最优化选择
Workflow Orchestration for On-Demand Real-Time Dynamically-Adaptive Systems	
Level 3-Application Services	
Grid testbed across the local institutions in Hong Kong	Science Portal
Advanced End-to-End Applications	TDDFT/Ninf-G
Advancing the treatment of bacterial infections	Teaching Radar through Advanced E-learning
Application Service Provision	Terascale Simulation Software
ATLAS Trigger/DAQ system	The eminerals portal
Aviation Safety Project	videoconferencing
BioGrid	video-mediated learning
Broadband Applications	video-teleconferencing facilities
Comprehensive Test Ban Treaty Organization	web service
adopting Grid technologies	Web-Based Distributed Simulation
distributed cognitions	差错数据分析
Earth Observing System	产品和流程开发
E-Kit	城市交通信息服务
e-learning	存档
Electron Micrograph Analysis	大气信息

eMinerals minigrid	大学课程在线
Enabling software	电子调度
Encyclopedia of Life	动态路况显示
Exerciser	动态路况最优选择
FacultyDevelopment.ca	多媒体访问
GADU/Gnare	仿真
GPS 终端器	仿真应用
Grass GIS	分布式诊断
Grid applications	高能物理
Grid middleware	个性化定制服务
Grid portal (optional or deployed in a later stage)	共享
grid web services	海量信息处理
GridFTP	航空制造
GridSAT - Boolean Satisfiability	互操作
GT3 评估	计算流体力学
ICECUBE Software	交通信息网格
Interfacing Interactive Analysis Tools	教育
Linked Environments for Atmospheric Discovery	决策支持
Management software	科学数据
mathmatics tools	空间科学
National FusionGrid	配置
NetLogger-grid2003	气象
OGSA	森林资源与林业生态工程
OGSA 评估	上海市交通局
Open Network Craft	生物信息
Particle Physics Data Grid	生物信息学
PLP Enabler	testbed
QM-MD/Ninf-G	通知服务
Rural Advanced Community of Learners	统一调度和管理
Savannah/Nimrod	图像处理
SchoolNet Caching	新药研发
scientific simulation codes	信息获取
site management tools	油气地震勘探应用
SnB	知识管理
software packaging tools	注释
System monitoring	资源环境
Improving image-guided surgery, minimizing invasive animal and human testing	
Interactive Multimedia Learning Systems for Mathematics on Broadband Networks	

Level 3-Grid Computational Services	
adaptive Grid services	MOP
adaptive grid techniques	Nimrod/G
Adaptive mesh refinement	Ninf-G
adaptive services	Performance Prediction
Common Data Services Architecture	Portability Manager
computational power grid	pyGlobus/CogKit
coordinate transformation technique	real-time distributed computing
d0grid	SAMGrid
data cyberinfrastructure	Stretched Grid Model Intercomparison Project
data grid	stretched grid
Distributed Processing Environment	VO Management Service eXtension
EcoGrid	生物信息学网格
grid adaptation	数值天气预报系统
Grid Canada	未来气象数值预报业务与协同研究开发
Grid Computing Portals	新药研发网络
Internet-D	中国气象局计算资源与研究人员
IPG Job Manager	中国气象应用网格
Level 3-Domain/ Task Ontologies and Metadata	
Automatic metadata capture	本体管理
Ecological Metadata Language	本体推理和查询
Hemp Hybrid Event store Metadata Prototype	本体应用
Knowledge Management	本体语言
Metadata Catalog	服务本体
Metascheduler for BioGrid	符号数学
Morpho	机构本体
natural language	领域本体
Optimisation Ontology	任务本体
User Ontology	术语学
Vocabulary	数值数学
本体表示	应用本体
本体服务	语义网络
本体服务的创建	元数据
本体服务的提供	元数据表示
本体工具	元数据分类
Level 3-Problem Solving Environments	
Advanced Broadband Enabled Learning	
Broadband Enabled Lifelong Learning Environment	

Barrier-free Broadband Learning Environments and The Inclusive Learning Exchange
CFD Solvers
E-learning in a Team-based Health-care Environment
InterNet Control and Inference Tools at the Edge
multiple mesh and discretization strategies
national/international leadership learning environment
New Integrated Multimedia Learning Environment for Internet Training
partial differential equations Simulations
SavoirNet
Terascale Simulation
基于网格的问题求解

附录G 参与国际合作项目3个以上的29个国家之间的合作矩阵

国别	德国	英国	美国	意大利	法国	荷兰	西班牙	波兰	奥地利	希腊	挪威	瑞典	瑞士	中国	澳大利亚	加拿大	爱尔兰	比利时	芬兰	韩国	以色列	丹麦	捷克	日本	葡萄牙	斯洛伐克	匈牙利	印度	
德国	30	17	3	6	7	9	8	8	5	6	7	5	5	3	3	2	5	5	4	2	3	0	4	1	3	1	1	2	
英国	17	28	3	10	10	7	10	6	5	6	5	5	7	3	1	3	4	5	2	2	3	1	4	1	2	2	3	2	
美国	3	3	21	2	3	3	1	2	1	0	2	1	3	2	4	4	1	0	1	3	0	0	2	1	1	1	0	3	
意大利	6	10	2	18	5	5	4	4	4	5	4	2	3	2	1	2	2	1	2	1	2	3	1	1	1	1	1	1	
法国	7	10	3	5	16	5	6	1	3	2	3	3	2	0	1	2	3	1	1	2	1	1	1	0	1	1	0	1	
荷兰	9	7	3	5	5	13	3	3	3	4	1	1	4	3	1	3	2	1	1	3	0	2	2	1	1	1	0	1	
西班牙	8	10	1	4	6	3	13	1	4	3	3	3	5	1	1	1	3	3	1	1	1	0	1	1	3	2	2	1	
波兰	8	6	2	4	1	3	1	11	2	2	3	2	3	2	2	2	2	0	3	2	0	1	2	1	1	2	0	2	
奥地利	5	5	1	4	3	3	4	2	9	2	2	1	4	1	2	1	5	3	2	1	2	0	1	1	1	2	0	1	
希腊	6	6	0	5	2	4	3	2	9	1	0	1	0	0	0	0	1	1	1	0	1	0	2	0	0	1	0	0	
挪威	7	5	2	4	3	1	3	3	2	1	8	5	2	2	2	2	2	3	4	1	0	1	1	1	1	1	0	1	
瑞典	5	5	1	2	3	1	3	2	1	0	5	8	1	1	2	1	1	2	3	1	1	2	1	1	1	2	1	0	1
瑞士	5	7	3	3	3	4	5	3	4	1	2	1	8	2	0	2	3	3	1	2	1	1	1	1	1	1	0	2	
中国	3	3	2	2	2	3	1	2	1	0	2	1	2	8	2	2	1	0	1	4	0	0	2	2	1	1	0	3	
澳大利亚	3	1	4	1	0	1	1	2	2	0	2	2	0	2	7	1	2	0	2	2	0	0	1	3	1	1	0	2	
加拿大	2	3	4	2	1	3	1	2	1	0	2	1	2	2	1	6	1	0	1	1	0	1	1	2	1	1	0	1	
爱尔兰	5	5	4	1	2	2	3	2	5	1	2	1	3	1	2	1	5	2	2	1	1	0	1	1	1	0	0	0	
比利时	5	5	0	1	3	1	3	0	3	1	3	2	3	0	0	0	2	5	0	0	1	0	0	0	0	0	0	0	
芬兰	4	2	1	2	1	1	1	3	2	1	4	3	1	1	2	1	2	0	5	1	0	1	1	1	1	1	0	1	
韩国	2	2	3	1	1	3	1	2	1	0	1	1	2	4	2	1	1	0	1	5	0	0	1	3	1	1	0	3	
以色列	3	3	0	2	2	0	1	0	2	1	0	1	1	0	0	0	1	1	0	0	5	0	0	0	0	0	0	0	
丹麦	0	1	0	3	1	2	0	1	0	0	1	2	1	0	0	1	0	0	1	0	0	4	0	0	0	0	0	0	
捷克	4	4	2	1	1	2	1	2	1	2	1	1	1	2	1	1	1	0	1	1	0	0	4	1	1	1	0	2	
日本	1	1	1	1	0	2	1	1	1	0	1	1	1	2	3	2	1	0	1	3	0	0	1	4	1	1	0	2	
葡萄牙	3	2	1	1	1	1	3	1	1	0	1	2	1	1	1	1	1	0	1	1	0	0	1	1	3	2	0	1	
斯洛伐克	1	2	1	1	1	1	2	2	2	1	1	1	1	1	1	1	1	0	0	1	1	0	0	1	2	3	0	1	
匈牙利	1	3	0	1	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0	
印度	2	2	3	1	1	1	1	2	1	0	1	1	2	3	2	1	0	0	1	3	0	0	2	2	1	1	0	3	

附录H 调研e-Science项目的主要学科及应用领域

生物科学		医药卫生		工业技术
基因组学		e-Health		材料科学
神经系统		临床医学		材料科学计算和工程模拟
生命科学		临床癌症信息学		材料设计
生命多样性		临床治疗		工程学
生态信息学		外科手术		宇宙空间和组织工程学
生态学		医学		科学与工程研究
生物分子学		脑科学		工程与物理
生物工程		生物医学		传统工业领域
生物工艺学		生物医学信息		发动机模拟
生物化学		医疗保健		工业制造
生物技术		医学成像		飞机制造
生物科学数据库		医学模型与仿真		汽车制造
生物物理		医学图像处理		润滑油制造的参数模型
生物信息学		医学图像与信号研究		洪水预测和物理学分析
生物学图片处理		医学信息学		洪灾预防
生物与环境研究		医药		洪灾预报模拟
细胞分析		制药领域内的产品设计		基础能源科学
遗传学		生物医药		
		药物设计		
		药物制造		
物理学		地球科学		天文学
电子学		地理		空间科学
高能物理		地球观测		空间气象
高能加速器		地球物理学		空间信息
高能物理实验		地球系统科学		超新星研究
高能与核试验		地震学		太阳物理研究
高能与核物理		地质学		天体物理学
磁聚变		石油工程		宇宙
核聚变学		大气		

续表

力学		大气层气候模型		物理学（续）
计算流体力学		大气和生物信息		等离子微观湍流研究
多材料流体力学		大气物理学		等离子物理学
粒子物理		气象领域内服务提供		纳米技术
粒子技术		气象学		
粒子物理计算		气候研究		
能源粒子物理		天气预报		
包括光谱		海洋学		
超大强子对撞机探测器		水文学		
紧凑子μ子螺线管探测器相关物理试验				

附录I 英国e-Science评估指标

I.1 英国e-Science核心计划的e-Science项目评估指标

针对英国 e-Science 当前的发展,英国,e-Science 指导委员会推出了针对 e-Science 核心计划的 17 项评估指标。

- ① 英国 e-Science 要求的经常性调查 (regular, survey) (定性评估)
- ② 短期、中期、长期项目的数量与质量 (定量与定性评估结合)
- ③ 开放中间件基础设施研究所 (OMII) 的影响力 (定性评估)
- ④ 英国 e-Science 研究院召开的会议数 (定量评估)
- ⑤ e-Science 研究院用户群的增长情况 (定量评估)
- ⑥ 安全政策的贯彻落实 (定性评估)
- ⑦ 年度专题会议、年度股东大会、镇代表年度双边会议的出席情况 (定量评估)
- ⑧ 数据掌管中心的影响力 (定性评估)
- ⑨ 2004 年秋投入运营的、基于 GT2 的工作网格 (定性评估)
- ⑩ 学术界支持机制的应用情况 (定性评估)
- ⑪ NSF, TeraGrid 的应用情况 (定性评估)
- ⑫ 开放中间件基础设施研究所国际活动范围 (定性评估)
- ⑬ 英国在 e-Science 标准制定活动中的参与度和发挥的作用 (定量评估)
- ⑭ 国际合作项目数 (定量评估)
- ⑮ 工业应用网格试点项目数 (定量评估)
- ⑯ 与主要用户的互动 (定量评估)
- ⑰ 企业界投资总额 (定量评估)

I.2 英国七大研究理事会的e-Science项目评估指标

(1) 英国生物技术与生物科学研究理事会 (BBSRC) 的 e-Science 评估指标

在 2000 年公开宣布的英国 e-Science 计划中,生物技术与生物科学研究理事会获得 800 万英镑的拨款用于建设 e-Science 项目。到 2002 年,该研究理事会又获得另外一笔高达 1100 万英镑的 e-Science 建设经费。到目前为止,该研究理事会已分两期资助 e-Science 项目。

第一期资助于 2002 年实施,共计启动五个项目;第二期资助于 2004 进行,同样启动了五个项目。由于该研究理事会强调“所有项目均以应用为导向”。因此,它所资助的 e-Science 项目均有生物领域密切相关。该研究理事会于 2004 年推出的 e-Science 项目评估报告设立的评估指标、评估机制明显带有生物科学的色彩。

生物技术与生物科学研究理事会对自身所承担的 e-Science 项目采取评估指标包括以下 16 项:

- ① 在分子生物学、细胞生物学和有机体生物学方面资助的项目数, (定量评估)
- ② 与投资者的互动 (engagement) (定量评估)
- ③ 以网格为实现基础的解决方案的数量与质量 (定量与定性评估相结合)
- ④ 软件与解决方法的用户范围 (定量评估)
- ⑤ 与企业界的合作情况 (定性评估)
- ⑥ BBSRC 所建立的网格支持中心的应用范围 (定性评估)
- ⑦ 参与培训的人员数量以及培训层次 (定量评估)
- ⑧ 面向用户与投资者的调查 (定性评估)
- ⑨ 利用国家 HPC (高性能计算) 的服务的情况 (定量评估)
- ⑩ 通过 BBSRC 所属的研究院传播 e-Science 的专业知识 (定量评估)
- ⑪ 对国际标准的影响 (定性评估)
- ⑫ 在国际 e-Science 项目中英国的参与程度 (定量评估)
- ⑬ 以国际合作方式开展的项目数 (定量评估)
- ⑭ 是否开展试点项目证明中英开展 e-Science 合作的可靠性 (定性评估)
- ⑮ 公司交互的数量 (定量评估)
- ⑯ 安全的 (外国) 对内投资额度 (定量评估)

(2) 中心实验室研究理事会 (CCLRC) 的 e-Science 项目评估指标

中心实验室研究理事会自身所承担的 e-Science 项目采取评估指标包括以下 11 项:

- ① 新开发的存取与使用机制被认可的程度 (定性评估)
- ② 科学成果 (定性评估)
- ③ 基于协同完成的项目而建立的社区数 (定量评估)
- ④ 基于网格的解决方案的数量、质量和使用状况 (定量与定性评估相结合)
- ⑤ 被认可的实践指南数量 (定量评估)
- ⑥ 指南实施的评估效果 (短期效果、中期效果与长期效果) (定性评估)
- ⑦ CCLRC 的设施与服务在国际合作中的使用情况 (定性评估)
- ⑧ 已经归档的科研成果 (定量评估)
- ⑨ 在国家 HPC/HEC 服务中的参与程度 (定性评估)
- ⑩ 已经建立的战略伙伴关系的数量、价值与持续时间 (定量与定性评估相结合)
- ⑪ 商业合作的数量、质量与价值 (定量与定性评估相结合)

(3) 工程与物理科学研究理事会 (EPSRC) 的 e-Science 项目评估指标

工程与物理科学研究理事会自身所承担的 e-Science 项目采取评估指标包括以下 22 项:

- ① 项目倡议书的质量 (定性评估)
- ② 资助的应用项目数 (定量评估)

- ③ 研究团队的组成（定量与定性评估相结合）
 - ④ 面向 e-Science 计划、与计算机科学密切相关的项目数量与质量（定量与定性评估相结合）
 - ⑤ 由英国 e-Science 核心计划、联合信息系统委员会（JISC）以及其他资助机构等联合资助的项目数量（定量分析）
 - ⑥ 发表论文数（定量分析）
 - ⑦ 获批准的专利数（定量分析）
 - ⑧ 软件与方法论被使用的范围（定量分析）
 - ⑨ 符合开放标准和 OSS 的成果数（定量分析）
 - ⑩ EPSRC 召开的年度会议数（定量分析）
 - ⑪ 参与英国 e-Science 年度专题研讨会（AHM）和其他展示会（定量分析）
 - ⑫ 项目涉及的应用领域（定量分析）
 - ⑬ 面向 e-Science 的计算机科学发展的水平（定量分析）
 - ⑭ 学生参与情况（定量分析）
 - ⑮ 最好的项目数（定量分析）
 - ⑯ 参与国家高端计算服务（定性评估）
 - ⑰ 从 SRIF 资金中分配给 e-Science 基础设施的资金数量（定量分析）
 - ⑱ 各项目组参与国际计划、国际项目和国际活动的情况（定性评估）
 - ⑲ 国际合作水平（定性评估）
 - ⑳ 面向特定学科领域、具有国际领先地位的项目（定量分析）
 - ㉑ 企业界为各项目投稿的资金数量（定量分析）
 - ㉒ 参与项目建设的公司数量（定量分析）
 - （4）经济和社会学研究理事会（ESRC）e-Science 项目评估指标
- 英国经济和社会研究理事会对自身所承担的 e-Science 项目采取评估指标包括以下 14 项：
- ① 项目质量（定性评估）
 - ② e-Science 年度专题研讨会上对年度报告和项目汇报的评估而引起的影响（定性评估）
 - ③ 与项目资助者进行互动的性质与影响（定性评估）
 - ④ 基于网格的解决方案的质量与影响（定性评估）
 - ⑤ 对当前伦理、机密性和安全性方面的影响（定性评估）
 - ⑥ 项目的合作情况（定量评估）由英国 e-Science 核心计划、联合信息系统委员会（JISC）以及其他资助机构等联合资助的项目数量（定量评估）
 - ⑦ 演讲与参与的专题讨论会（定量评估）
 - ⑧ 科学成果（包括：开设的培训课程、文章、学生参与情况和国际合作水平）（定量评估）

⑨ 对方法论、学科与数据分析方面的影响（定性评估）

⑩ 对资助者的影响（定性评估）

⑪ 对国际标准的影响（定性评估）

⑫ 国际互动的性质与影响（定性评估）

⑬ 安全的（外国）对内投资的水平与影响（定量评估）

⑭ 期望从示范性项目中获到的培训资料（定性评估）

（5）医学研究理事会（MRC）e-Science 项目评估指标

医学研究理事会对自身所承担的 e-Science 项目采取评估指标包括以下 12 项：

① 从 e-Science 年度专题研讨会上的年度报告和项目汇报反映出来的项目质量与影响（定量评估）

② 与项目资助者进行互动的性质与影响（定量评估）

③ 基于网络的解决方案的质量与影响（定性评估）

④ 对当前伦理、机密性和安全性方面的影响（定性评估）

⑤ 协同合作的试点项目数（定量评估）

⑥ 接受培训的人员学历层次以及对项目参与人员的影响（定量评估）

⑦ 对方法论、学科与数据分析方面的影响（定性评估）

⑧ 对项目资助机构的影响（DH、DTI、UKERNA）（定性评估）

⑨ 对国际标准的影响（定性评估）

⑩ 国际互动的性质与影响（定量评估）

⑪ 与公司企业互动的性质与影响（定量评估）

⑫ 安全的（外国）对内投资的水平与影响（定量评估）

（6）自然环境研究理事会（NERC）e-Science 项目评估指标

自然环境研究理事会对自身所承担的 e-Science 项目采取评估指标包括以下 18 项：

① 用的数量与质量（定量与定性评估相结合）

② 奖学金的数量（定量评估）

③ 科学成果——案例研究（定性评估）

④ 项目组互动交流的数量（定量评估）

⑤ 在 PI, Col 研究院中 RAE 排名（定量评估）

⑥ 发表的论文数（定量评估）

⑦ 接受培训的人员数量与学历层次（定量评估）

⑧ NERC 与企业界、学术界交流的情况（定量评估）

⑨ 参与培训与专题研讨会的情况（定量评估）

⑩ Pis 与 Cols——与 5*部门相关的项目数（定量评估）

⑪ 设施的可使用性（定性评估）

⑫ NERC 所属的研究中心、著名大学中 e-Science 专门知识的提供情况（定性评估）

- ⑬ 国际伙伴关系的数量和重要性（定量评估）
- ⑭ 国际伙伴关系的数量、价值与重要性（定量评估）
- ⑮ 通过国际合作而完成的出版物（定量评估）
- ⑯ 商业合作关系的数量和价值（定量评估）
- ⑰ 与企业界、政府和用户的互动与交流（定性评估）
- ⑱ 承担 SBRI 的情况（定性评估）

（7）粒子物理研究理事会（PPARC）e-Science 项目评估指标

粒子物理研究理事会对自身所承担的 e-Science 项目采取评估指标包括以下 11 项：

- ① 通过虚拟组织基础设施进行存取的英国重要数据中心的数量（定量评估）
- ② 在 2005 年 LHC 计算系统技术设计报告中英国对 CERN 的贡献程度（定性评估）
- ③ 使用虚拟组织工具包的用户库规模（定量评估）
- ④ 使用 GridPP 的用户库规模（定量评估）
- ⑤ GridPP 与英国 e-Science 网格之间公共基础设施组件的数量（定量评估）
- ⑥ 由 PPARC 所资助的、在其他学科中从事 e-Science 研究的人员数量（定量评估）
- ⑦ 从 LHC 实验中获取第一批数据的准备情况、英国学术界对 GridPP 认可程度（定性评估）
- ⑧ 将欧洲网格的发展与英国战略计划进行比较、英国科研人员在欧洲项目中的参与情况与领导地位（定性评估）
- ⑨ 新任务与新设施的数量（能够制造可通过虚拟组织被研究人员存取的数据）（定量评估）
- ⑩ GridPP 项目中国际伙伴者的数量及覆盖范围、其他国家使用虚拟组织系统交流数据的发展程度（定量评估）
- ⑪ 在国际合作项目中拥有重要学术地位的英国研究人员数（定量评估）

附录J CORE、七大研究理事会e-Science项目 评估指标

	C O R E	B B S R C	C C L R C	E P S R C	E S R C	M R C	N E R C	P P A R C
经常性调查	√	√						
项目的数量与质量	√	√		√	√	√		
OMII 的影响力	√							
会议数、专题研讨会等	√			√	√		√	
安全政策	√							
数据掌管中心的影响力	√							
GT2 的工作网格	√							
学术界支持	√							√
TeraGrid 的应用情况	√							
与相关标准的关系	√	√		√	√	√		
培训相关		√			√	√	√	
工业应用网格试点项目数	√	√						
用户、企业的合作与交流	√	√		√		√	√	
国际合作与交流	√	√	√	√	√	√	√	√
投资者的交流		√			√	√		
网格为实现基础的解决方案的数量与质量		√	√		√	√		
用户规模								√
成果使用范围	√	√		√			√	√
投资额	√	√		√	√	√		
新开发的存取与使用机制被认可的程度			√					
科学成果（论文数、专利数）			√	√	√		√	
基于协同完成的项目而建立的社区数			√					
被认可的实践指南数量			√					
指南实施的评估效果			√					
项目倡议书的质量				√				
研究团队的组成				√				

续表

	C O R E	B B S R C	C C L R C	E P S R C	E S R C	M R C	N E R C	P P A R C
面向 e-Science 计划、与计算机科学密切相关的项目数量与质量				√				
学生参与情况				√				
参与国家高端计算服务		√	√	√				√
面向特定学科领域、具有国际领先地位的项目				√				
伦理、机密性和安全性					√	√		
对方法论、学科与数据分析方面的影响					√	√		
对资助者的影响					√	√		
协同合作的试点项目数						√		
在 PI, Col 研究院中 RAE 排名							√	
Pis 与 Cols——与 5*部门相关的项目数							√	
设施的可使用性							√	
通过国际合作而完成的出版物							√	
承担 SBRI 的情况							√	
虚拟组织基础设施进行存取的英国重要数据中心的数量								√
使用虚拟组织工具包的用户库规模								√
GridPP 与英国 e-Science 网格之间公共基础设施组件的数量								√
由 PPARC 所资助的、在其他学科中从事 e-Science 研究的人员数量								√
将欧洲网格的发展与英国战略计划进行比较、英国科研人员在欧洲项目中的参与情况与领导地位								√
新任务与新设施的数量								√
在国际合作项目中拥有重要学术地位的英国研究人员数								√

附录K 本书所用缩略语和中英文对照表

英文缩写	英文全称	中文全称
BMBF	Bundesministerium für Bildung und Forschung	德国联邦教育与研究部
CERN	European Organisation for Nuclear Research	欧洲原子能机构
CLRC	Central Laboratory of the Research Councils	英国研究理事会中心实验室
CNR	Consiglio Nazionale delle Ricerche	意大利国家研究理事会
Datamat	DATAMAT S.p.A.	DataMat 公司
DOE	Department of Energy	美国能源部
EGEE	The Enabling Grids for E-science in Europe	欧洲网格计划
EPSRC	The Engineering and Physical Sciences Research Council	英国工程与物理科学研究理事会
ERA	European Research Area	欧洲研究领域
FLE	Fujitsu Laboratories of Europe	富士通欧洲实验室
FNS	Fonds National de la Science	法国研究部国家科学基金
Fraunhofer	Fraunhofer Institute	Fraunhofer 研究所
FZJ	Forschungszentrum Jülich	德国尤利希研究中心
HLRS	Universität Stuttgart, High Performance Computing Centre Stuttgart	斯图加特高性能计算中心
INRIA	Institut National de Recherche en Informatique	法国国立计算机科学与控制研究所
ICM	Interdisciplinary Centre for Mathematical and Computational Modelling, University of Warsaw	华沙大学数学与计算机建模交叉研究中心
IT-Inovation	The University of Southampton, IT Innovation Centre	南安普顿大学 IT 创新中心
ITR	Information Technology Research for National Priorities	美国信息技术研究
MIUR	Ministero dell'Istruzione, dell'Università e della Ricerca	意大利教育大学和研究部
NASA	National Aeronautics and Space Administration	美国航太总署
NMI	NSF Middleware Initiative	美国科学基金会中国件首倡计划
NSF	National Science Foundation	美国科学基金会
NTUA	the Institute of Communication and Computer Systems at the National Technical University of Athens	雅典国立技术大学通信和计算机系统研究所
PRAGMA	Pacific Rim Applications and Grid Middleware Assembly	环太平洋网格应用与中间件联盟
RNRT	RESEAU NATIONAL DE RECHERCHE EN TELECOMMUNICATIONS	国家网络互联和电信网络
RNTL	Réseau National des Technologies Logicielles	国家软件工程网络
UCY	University of Cyprus, Dept. of Computer Science	塞浦路斯大学计算机科学系
UNICORE	Uniform Interface to Computing Resources	面向计算资源的统一界面
UMAN	The Victoria University of Manchester	曼切斯特维多利亚大学
UvA	Universiteit van Amsterdam, Faculty of Science	阿姆斯特丹大学科学系
ZIB	Konrad-Zuse-Zentrum für Informationstechnik Berlin	柏林 Zuse 研究所